JUMS

Volume 1, Issue 1, June 2016

JUNIOR MANAGEMENT SCIENCE

Advisory Editorial Board:

DOMINIK VAN AAKEN
FREDERIK AHLEMANN
CHRISTOPH BODE
ROLF BRÜHL
LEONHARD DOBUSCH
RALF ELSAS
DAVID FLORYSIAK
GUNTHER FRIEDL
WOLFGANG H. GÜTTEL
CHRISTIAN HOFMANN
KATJA HUTTER
STEPHAN KAISER
ALFRED KIESER
NATALIA KLIEWER
DODO ZU KNYPHAUSEN-AUFSEß
SABINE T. KÖSZEGI
ARJAN KOZICA
TOBIAS KRETSCHMER
HANS-ULRICH KÜPPER
REINER LEIDL
ANTON MEYER
GORDON MÜLLER-SEITZ
GÜNTER MÜLLER-STEWENS
BURKHARD PEDELL
ARNOLD PICOT
MARCEL PROKOPCZUK
TANJA RABL
SASCHA RAITHEL
ASTRID REICHEL
KATJA ROST
MARKO SARSTEDT
DEBORAH SCHANZ
ANDREAS G. SCHERER
STEFAN SCHMID
UTE SCHMIEL
PHILIPP SCHRECK
GEORG SCHREYÖGG
LARS SCHWEIZER
DAVID SEIDL
ANDREAS SUCHANEK
ANJA TUSCHKE
SABINE URNIK
STEPHAN WAGNER
ISABELL M. WELPE
CLAUDIA B. WÖHLE
THOMAS WRONA
THOMAS ZWICK

Alexander Wahl, Variance Risk Premia 1

Lennart Wendland, A Common Ground in Guerilla Marketing – State of Research and Further Research Opportunities 34

Friederike E. Rhein, Rückzug auf Raten – Management strategischen Wandels: Eine retrospektive Längsschnittstudie am Beispiel der ChemCo. 60

Nadine Eichhorn, Relative Performance Information and Financial Incentives in Multidimensional Task Settings – A Conceptual and Experimental Analysis of Effects on Performance and Attention towards Task Dimensions 100

Nina Gusenleitner, Practical Implications of the Ambidexterity Concepts 138

Nicolas Wüthrich, Inkommensurabilität – Implikationen für die evolutionäre Organisationstheorie unter begriffstheoretischer und epistemologischer Perspektive 188

Sonja A.-M. Hammel, Effektive Prävention ethisch-moralisch bedingter Unternehmenskrisen – Eine fallbasierte Untersuchung in Deutschland und den USA 216

Johannes Weindl, Potenziale von Batteriespeichersystemen zur Regelenergiebereitstellung 275

Published by Junior Management Science

Vorwort

Die vorliegende, erste Ausgabe von Junior Management Science (JUMS) ist in vielerlei Hinsicht besonders. Insbesondere ist sie ein Zeichen gelungener hochschulübergreifender Kooperation. So finden sich im Advisory Editorial Board über 25 Universitäten aus Deutschland, Österreich und der Schweiz wieder, ohne die die vorliegende Zeitschrift nicht hätte realisiert werden können. Diese Kooperation basiert auf einem gemeinsamen Verständnis, welches Forschung und Lehre als Einheit betrachtet und studentische Leistungen wertschätzt.

Die Zielsetzung von JUMS ist es, herausragende studentische Leistungen auf dem Gebiet der Betriebswirtschaftslehre zu identifizieren und zu würdigen. Gemäß dieser Zielsetzung ist JUMS offen für die gesamte inhaltliche und methodische Breite der BWL. Arbeiten zu klassischen Bereichen wie der Finanzierung, dem Personalwesen und der Unternehmensrechnung sind genauso willkommen wie zu neueren Gebieten, wie beispielsweise der Unternehmensethik und des Neuromarketings. Der theoretische Hintergrund kann dabei in der Ökonomik, Psychologie, Soziologie, aber auch in anderen Disziplinen liegen. Auch gibt es keine vorgeschriebene methodische Ausrichtung der Arbeiten.

In JUMS soll sich die Vielfalt der Betriebswirtschaftslehre widerspiegeln, was sich auch an den Inhalten dieser ersten Ausgabe ablesen lässt: Unter den Beiträgen finden sich quantitative, qualitative und konzeptionelle Abschlussarbeiten, die unter anderem Problemstellungen des Marketings, der Unternehmensrechnung, der Organisation bis hin zur Unternehmensethik behandeln. So unterschiedlich die hier versammelten Arbeiten hinsichtlich der zugrunde liegenden Methodik und disziplinären Verortung sind, gemein ist allen, dass sie – gemessen am aktuellen State of the Art – einen Erkenntnisfortschritt aufweisen.

Die in dieser ersten Ausgabe veröffentlichten Arbeiten basieren auf einer Empfehlung jeweils eines Mitglieds des Advisory Editorial Boards. Dies ist hervorzuheben, da wir schon in der nächsten Ausgabe die Artikel nach einem doppelt-blinden Begutachtungsprozess auswählen werden.

Durch den exklusiven Fokus auf die Veröffentlichung wissenschaftlich wertvoller Bachelor- und Masterarbeiten dient JUMS mehreren gesellschaftlichen Gruppen. WissenschaftlerInnen zum Beispiel erhalten einen Einblick in aktuelle Themenstellungen und Forschungsmethoden anderer Institute. Unternehmen und andere an betriebswirtschaftlicher Forschung Interessierte erhalten Zugang zu exzellentem wissenschaftlichen Wissen. JUMS veröffentlicht Open Access, was einen freien Zugang zu den Arbeiten für jeden sicherstellt. Das Studium wird weitgehend via staatlicher Finanzierung durch die Gesellschaft ermöglicht. Dass die besten Abschlussarbeiten der Gesellschaft dann frei zur Verfügung gestellt werden, liegt nahe.

Der Nutzen für die Studierenden ist für uns das zentrale Merkmal von JUMS. Zum einen erhalten Autoren durch JUMS eine frei zugängliche Plattform zur Verbreitung Ihrer Abschlussarbeit. Ihre Arbeiten sind somit zitierfähig, womit die erarbeiteten Erkenntnisse eine Chance zur Verbreitung haben. Zum anderen weist die Veröffentlichung der besten Abschlussarbeiten eine Vorbildfunktion für aktuell Studierende auf. JUMS schafft Transparenz bei Abschlussarbeiten: Nun ist es jedem Studierenden möglich, zu sehen, was WissenschaftlerInnen hochschulübergreifend als hervorragende Abschlussarbeit bewerten. Damit können sich aktuell Studierende bei der Erstellung Ihrer eigenen Abschlussarbeit an den in JUMS veröffentlichten Arbeiten orientieren.

Und nun wünschen wir Ihnen viel Spaß und neue Einsichten bei dem Studium der ersten Ausgabe von Junior Management Science.

Dominik van Aaken, David Florysiak und Sebastian Waic

Junior Management Science 1 (2016) 1-300

Junior Management Science

journal homepage: www.jums.academy

Variance Risk Premia

Alexander Wahl
Ludwig-Maximilians-Universität München

Abstract

Using a relatively model-free approach to extract the risk-neutral expected variance from an extensive set of traded options on 29 single stocks and eight stock indices, I derive the variance risk premium defined as the difference between the actually realized variance and the expected variance under the risk-neutral measure. The analysis reveals that variance risk premia are persistently negative for the majority of underlyings and show a clear link to the underlying's exposure to systematic market variance. Moreover, I find that both the risk associated with continuous as well as discontinuous price movements contribute to observed variance risk premia.

Keywords: Variance risk premium, Volatility premium, Jumps, Risk-neutral, Model-free

1. Introduction

1.1. Problem outline

Even though the option pricing model invented by Black and Scholes (1973) is probably the most famous of its kind and often used in practice (albeit with modifications), it is well known that particularly the assumption of a constant return variance is not fulfilled in practice and volatility fluctuates over time. By now a plethora of studies (e.g. Bakshi and Kapadia (2003), Coval and Shumway (2001), Carr and Wu (2009)) document evidence that certain option positions earn average returns significantly different from zero although these positions are constructed to be insensitive to fluctuations in the price of the underlying. Indeed, this return pattern is most often interpreted as evidence that investors price the risk associated with stochastic volatility, which gives rise to the so-called variance risk premium.

Over the years, a veritable market for volatility products has emerged and different parties such as for example hedge funds (e.g. Bondarenko (2004)) engage in volatility selling strategies or are otherwise exposed to the risk associated with changes in volatility.

With regard to this situation, it is important to know precisely whether return variance constitutes an independent risk factor that is priced by the market or whether the returns on financial instruments and strategies that are targeted at capturing the variance risk premium can be explained by other means. For one thing, a profound knowledge about the risks to which an asset offers exposure is necessary in the context of asset pricing to determine expected returns and asset prices that are commensurate to their risk exposure. Furthermore, this knowledge is of equal importance for related applications such as performance evaluations and attributions of institutional investors who engage in volatility related strategies and for risk management purposes since an assessment of the risks to which one is exposed arguably allows more efficient and successful hedging.

With reference to this, the objectives of this work are twofold. The first objective is to extract the variance risk premium from a set of traded options and to quantify its magnitude. Despite the fact that this is an approach taken by several other studies before (e.g. Bakshi and Kapadia (2003)), doing so can nevertheless provide interesting insights. Volatility related strategies are commonly regarded as extremely profitable, allowing to obtain Sharpe ratios that are substantially higher than simple equity investments (e.g. Carr and Wu (2009)). However, such strategies are often characterized by relatively frequent (small) gains but occasional losses of excessive magnitude (Ilmanen (2012)). Since unprecedented levels of volatility emanated during the period around and following the recent financial crisis and severely affected the returns on instruments whose value depends on return volatility, the dataset applied in this work offers interesting insights about how such strategies performed during this period and how average returns are affected over the long-run.

DOI: http://dx.doi.org/10.5282/jums/v1i1pp1-33

The second objective is to assess whether the extracted variance risk premium indeed represents compensation for the risk associated with a fluctuating return variance or whether the returns on such strategies can be explained by commonly applied asset pricing models.

1.2. Course of examination

For this work, I use an extensive set of options data on three stock indices from the United States (US) and five from Europe as well as on 29 single names from the US over the period from January 1996 until August 2015. In order to quantify the variance risk premium on each of these underlyings, I follow the procedure proposed by Carr and Wu (2009) and use the notion of a variance swap contract which is effectively a forward contract that pays the difference between the actually realized return variance over a predefined time period and a variance strike. The variance strike in this contract equals the expected risk-neutral variance and can be synthesized from a set of European options using a relatively model-free approach. Even though the estimate obtained with this method is subject to an approximation error when the asset price process is not purely continuous, it offers the advantage of a theoretical basis and does not impose further structural restrictions on the asset price process, which would inevitably be the case if one tried to calibrate a certain model to real data in order to derive the variance risk premium.

If the market does not require a non-zero risk premium that is embedded in the prices of traded options, the variance strike should equal the ex-post realized variance on average. As a consequence, it is possible to quantify the average variance risk premium as the time series average of the difference between the ex-post realized variance and the risk-neutral expected variance over the corresponding time period, i.e. the variance strike, or alternatively as the average return on a variance swap contract.

Since the results suggest that variance risk premia are indeed significantly different from zero and negative for all indices and the majority of single names, but differ substantially in their absolute magnitude, the analysis is continued with an attempt to link the variance risk premium, or variance swap returns, to the underlying's exposure to systematic variance. For this purpose, I compute a variance beta that measures the covariation between the return variance of a proxy for the market portfolio and that of the underlying under consideration. Variance swap returns are then regressed on this variance beta. Because this analysis reveals that the variance beta can only explain variance swap returns for the US indices but not for single stocks, it is further examined whether variance swap returns can be explained by a systematic variance risk factor that is proxied by the return on a variance swap with the S&P 500 as underlying. The underlying rationale for this course of action is that different studies find evidence that idiosyncratic return variances have a tendency to move together and possibly exhibit a common factor structure (Herskovic et al. (2014)) to which a diversified index, however, should not be exposed. Investors may require compensation for the risk associated with common movements in idiosyncratic return variances and this compensation may contribute to the observed variance risk premium (Schürhoff and Ziegler (2011), Gourier (2015)).

In order to further investigate whether return variance constitutes an independently priced risk factor, it is examined whether the classical Capital Asset Pricing Model (CAPM) or the Fama and French (1993) three-factor model are able to explain variance swap returns. Even though both models generate significantly negative regression betas with respect to market excess returns that are consistent with the commonly observed negative correlation between equity returns and return variance (Glosten et al. (1993)), regression alphas are mostly negative and significantly different from zero, which is suggestive of one or more additional priced risk factors.

Because the total return variance reflects the continuous, i.e. pathwise variation as well as discontinuous price movements or jumps, I attempt to examine whether the risk associated with each of these types of price variation separately commands compensation that contributes to the observed variance risk premium. For this purpose I construct two risk-factor-mimicking portfolios from traded options that are targeted at offering exposure to one risk while being relatively unaffected by the other. Since the validity of any conclusion about the extent to which jump risk is priced crucially depends on the ability of the jump-risk-mimicking portfolio to reliably capture the exposure to jumps, it is tested whether positive returns on this factor coincide with jumps detected by the non-parametric jump detection test of Lee and Mykland (2008), which appears to be the case. Even though both constructed risk factors significantly contribute to explaining variance swap returns and continue to be significant when commonly used risk factors are included as control variables, abnormal returns remain mostly negative and significantly different from zero. Due to this, I consider the effect of the chosen specification of variance swap returns. In the initial setting, regressions are performed using continuously compounded variance swap returns to account for the substantial skewness and kurtosis in the distribution of raw returns. However, this has the drawback of shifting the mean return further into the negative domain, which may contribute to the persistently negative alphas. Due to this, robustness tests with raw variance swap returns instead of continuously compounded returns are performed and suggest that the specification has a certain impact on results but abnormal returns, especially for the US indices, often remain significant.

Finally, motivated by a relatively recent study by Drechsler (2013) who finds that model uncertainty can help substantially to explain the magnitude of the observed variance risk premium, I examine the relation between a proxy for model uncertainty and variance swap returns.

2. Literature review

For one thing, this work is related to prior studies that examine whether a variance risk premium exists and in particular to studies that examine the existence of the variance risk

premium not only for stock indices but also for single stocks. One of the earlier studies that provide empirical evidence for the existence of a negative variance risk premium in index options is the work by Bakshi and Kapadia (2003) in which a long call option is dynamically hedged with the underlying. Since this strategy is effectively market-neutral and should therefore only be exposed to volatility risk, Bakshi and Kapadia (2003) interpret the negative average returns on the strategy as evidence for the existence of a negative volatility risk premium. While the outlined approach allows to infer the variance risk premium rather indirectly, Carr and Wu (2009) are among the first who use a model-free approach to synthesize variance swap rates from a sample of traded options on five US stock indices and 35 individual stocks in order to examine the existence and dynamics of the variance risk premium. This model-free approach allows a direct quantification of the variance risk premium which is defined as the difference between the risk-neutral expected and subsequently realized variance. Their results suggest that average variance risk premia are negative and substantial for the five indices whereas premia for individual stocks exhibit considerable variation, are not always significant and can even be positive. They further examine whether observed variance risk premia are related to a systematic variance factor proxied by the return variance of the S&P 500. This analysis reveals that assets with higher exposure to the variance factor are associated with more negative variance risk premia, which leads them to conclude that investors dislike elevated levels of market volatility and are willing to accept negative average returns on variance swaps to insure against rising market volatility. Moreover, Carr and Wu (2009) find that frequently used asset pricing models such as the Capital Asset Pricing Model or the Fama and French (1993) model are unable to explain variance swap returns.

Driessen et al. (2009) use the same model-free approach to derive variance risk premia embedded in S&P 100 stock index options and options on all its individual constituents. Similar to Carr and Wu (2009), they find a significantly negative variance risk premium for the S&P 100 whereas the variance risk premiums on individual stocks are often zero or even positive. Since an index variance risk premium should reflect the variance premia of all its constituents, they conclude that a negative variance risk premium for the index together with non-existent or even positive variance risk premia on its constituents is only reconcilable with priced correlation risk. In order to empirically test this hypothesis, they implement a trading strategy that is targeted at capturing the correlation risk premium and earns significant abnormal returns. Based on their results, Driessen et al. (2009) also link the expensiveness of index options relative to individual stock options to the insurance against undesirable marketwide correlation increases that index options offer but individual options do not.

Schürhoff and Ziegler (2011) also synthesize variance swap rates to decompose total variance risk and examine the separate pricing of systematic and idiosyncratic variance risk for the S&P 100 and NASDAQ 100 index and all index constituents. Consistent with previous findings, their results suggest that systematic variance carries a negative risk premium. Moreover, they also find that common idiosyncratic variance risk, i.e. the risk of comovements in the variances of idiosyncratic stock returns, carries a risk premium which is, on average, positive. Since the total variance risk premium is the sum of the two components, Schürhoff and Ziegler (2011) argue that the non-existent variance risk premium on individual stocks documented by Driessen et al. (2009) is due to the fact that the two components offset each other for S&P 100 constituents. They further attribute the positive risk premium on common idiosyncratic variance risk to financial intermediaries that are net-long in options on individual stocks. Since these intermediaries are not able to perfectly hedge options and are exposed to the associated idiosyncratic variance risk, they require compensation for upholding this position. Moreover, with regard to the results by Driessen et al. (2009), Schürhoff and Ziegler (2011) argue that the return on dispersion trading strategies that target to capture the correlation risk premium does not represent compensation for the pure risk of marketwide increases in correlation but rather reflects a combination of risk premia on systematic variance and common idiosyncratic variance risk.

In a related study, Gourier (2015) introduces an affine jump-diffusion model that considers the factor structure of asset returns as well as that of idiosyncratic return variance, in which variance swap rates and the variance risk premium – using the same definition as Carr and Wu (2009) – can be derived in closed-form. Further analyses show that model-based variance swap rates match synthetic variance swap rates remarkably well. Gourier (2015) finds a negative variance risk premium for all stocks that rises in absolute magnitude when the time to maturity increases. In contrast to Schürhoff and Ziegler (2011), she further finds that idiosyncratic variance risk carries a negative risk premium whose contribution to the overall variance risk premium is substantial and amounts to 80% on average.

The presented selection of studies shows that the existence of a negative variance risk premium for stock indices is well documented whereas the results and in particular the sign of the variance risk premium for single stocks are more ambiguous.

Apart from studies that document the existence of a variance risk premium, this work is also related to studies that attempt to examine whether the wedge between risk-neutral expected and actually realized variances predominantly reflects the risk that arises from stochastic volatility or rather results from compensation for the risk of discontinuous price movements, i.e. price jumps. For instance, Todorov (2010) examines the dynamics of the variance risk premium over time and especially how it is related to price jumps. For this purpose he fits a general semiparametric stochastic volatility model to S&P 500 data and uses deep out-of-the-money and close-to-maturity options to estimate the risk-neutral tail jump intensity. The chosen approach allows him to infer the contribution of stochastic volatility and jumps to the ex-ante variance risk premium which he measures as the difference

between a model-based estimate for the future expected variance under the physical measure and the VIX index. He concludes that jumps play a crucial role in explaining the variance risk premium. Since it increases after the occurrence of price jumps and reverts only slowly to its long-run mean thereafter whereas the impact of jumps on future market dynamics is limited, he concludes that investors' perception of jump risk is time-varying.

In a related work, Bollerslev and Todorov (2011) employ extreme value theory together with high frequency and options data in a nonparametric setting to decompose the equity risk premium and the variance risk premium into two separate components for diffusive and jump risk. Their results suggest that on average approximately 5% (in absolute terms) of the equity risk premium and more than half of the variance risk premium represent compensation for jump tail events.

3. Theoretical foundation of the variance risk premium

For a negative risk premium, such as the observed variance risk premium, to persistently occur, two complementary forces that are outlined in the following must coincide. First, the party requiring the risk premium must be exposed to essentially non-diversifiable, i.e. systematic, or non-hedgeable risk since otherwise a risk premium would not be justified and competitors would underbid each other until the market price of the asset equals its no-arbitrage price. The existence of such a non-diversifiable or non-hedgeable risk is strongly linked to the assumed price process of the underlying. It is well known that if the price of the underlying is assumed to follow a Geometric Brownian motion with constant return variance, as in the model by Black and Scholes (1973), markets are essentially complete and options are redundant securities because any option can be perfectly replicated by holding the commensurate amount of the underlying and the risk-free bond. The crucial aspect of such a setting is that the option's systematic risk is fully accounted for by the price of the underlying used in the replicating portfolio (Hull (2009)). Thus, in a world such as that described by Black and Scholes (1973), any exposure to options can be hedged perfectly and, if there is a competitive market for options, option prices should be solely determined by no-arbitrage conditions without the explicit consideration of risk premia. If stochastic volatility or random jumps in the asset price process are introduced, however, there are two additional variables that can change randomly and affect the price of an option. Consequently, it is no longer possible to perfectly replicate an option's payoff and thus hedge the option through simple trading in the underlying and the risk-free bond. As proven by Bajeux-Besnainou and Rochet (1996), introducing stochastic volatility that fluctuates independently of the price of the underlying (in contrast to a setting where volatility is a deterministic function of the asset price (e.g. Dupire (1994))) in a continuous time setting makes a classical European option always a non-redundant security. Because neither stochastic volatility nor jumps are traded assets, the market is incomplete with respect to states in the world where these variables change (Staum (2007)). Technically, market incompleteness means that there is no unique equivalent martingale measure, or risk-neutral probability distribution, for which all discounted asset prices are martingales, but multiple such measures exist (Bajeux-Besnainou and Rochet (1996)). Due to the absence of assets that allow to trade these particular risks, it is also not possible to complement the underlying and the risk-free bond with an additional asset whose price exclusively depends on these variables and thereby replicate and hedge the option perfectly. As a consequence, the holder of an option position is left with the essentially non-hedgeable risk of random changes in the variance of the underlying or price jumps and associated changes in the value of her option. Thus, depending on whether these additional sources of risk are systematic and therefore correlated with aggregate consumption, they may command a risk premium. For instance, in the model proposed by Hull and White (1987)), stochastic volatility does not carry a non-zero risk premium because it is explicitly assumed to be uncorrelated with aggregate consumption. Likewise, for example Merton (1976) assumes that the source of price jumps is company- or industry-specific information so that the contribution of price jumps to stock returns is non-systematic. However, there is considerable empirical evidence that return variances of different stocks tend to move together (e.g. Andersen et al. (2001)) and correlations in changes of implied volatilities make it difficult to eliminate vega risk by simple diversification (Engle and Figlewski (2014)). Thus, it can reasonably be assumed that variance risk is systematic and difficult to diversify. As a consequence, options will be priced to reflect the required compensation for random changes in the volatility (e.g. Bates (2000)). Similarly, Ang and Chen (2002) examine how correlations differ for upside and downside moves and find that correlation asymmetries are more pronounced for extreme downward moves, which can be interpreted as an indirect indication that jump risk may be systematic, thus potentially justifying a jump risk premium.

Note that for example in the model by Heston (1993), an additional option that is already traded would complete the market (Staum (2007)) and theoretically allow to perfectly hedge the risk of random changes in return variance. However, even in this situation, a perfect hedge still requires that the applied option pricing model be correct and the correct input parameters be used to reliably estimate each option's sensitivity with respect to random changes in the return variance. This leads to the closely related practical problem of model uncertainty. In this context Broadie et al. (2009) argue that market makers in option markets may require a risk premium since the necessary estimation of parameters such as the spot volatility, long-run mean levels of volatility and volatility mean reversion parameters and the associated determination of hedge ratios in the presence of stochastic volatility and jumps is subject to considerable estimation risk and may lead to the effect that option market makers cannot hedge their option exposure perfectly. For instance, Green

and Figlewski (1999) analyze the effect of inaccurate volatility estimates on delta-hedged short positions in index options by comparing the performance using as input to the option pricing model the best historical estimate of volatility and actually realized volatility over the remaining life of the option. Their results show that model risk due to an inaccurate volatility estimate has economically substantial effects even when positions are delta-hedged daily. Thus, the pricing of this estimation risk by risk-averse market makers could also contribute to the observed difference between option-implied and realized variances.

Empirical evidence that option market makers may indeed find it difficult to hedge their inventory is given by Gârleanu et al. (2009) who argue that option prices should not be affected by demand pressures when competitive intermediaries can hedge their option positions perfectly. They formalize in a model the notion that risk-averse intermediaries require compensation for their inability to hedge their exposure and link the amount of unhedgeable risk to the net-position these market makers hold. Their empirical results suggest that option expensiveness, defined as the difference between the average of options' implied volatility and a volatility forecast is indeed related to the net-position market makers have in options. In particular, index options in which intermediaries are net-short so that increases in volatility are associated with losses to the intermediary are visibly more expensive than individual equity options in which intermediaries are net-long. Further, Fournier and Jacobs (2015) also assume that market makers face unhedgeable risks and find that an option market makers' increasing inventory exposure to market variance risk, which is measured as the aggregate Black and Scholes (1973) vega, is associated with a significantly more negative variance risk premium (using the same definition of the variance risk premium as applied here). In this context, the results of Gârleanu et al. (2009) and Fournier and Jacobs (2015) offer an interesting explanation for the differential magnitude of variance risk premia in equity index options and options on individual stocks (e.g. Carr and Wu (2009), Driessen et al. (2009)) with regard to the net-position held by market makers in these contracts. Note, however, that the cited studies take the demand for options as exogenously given. Thus, they offer a potential explanation for why market makers may demand a variance risk premium but leave open the question why investors should be willing to pay it. This leads to the second prerequisite for the persistent existence of a variance risk premium.

While market makers seem to be exposed to unhedgeable systematic risk which justifies a risk premium, the second prerequisite for its persistent existence is that the assets on which an investor pays the premium, namely (index) options and variance swaps, offer insurance against an undesirable state of the world. If this was not the case, investors would not be willing to pay the substantial negative premium that is often found. A risk-averse investor is usually characterized by marginal utility that decreases with wealth. Due to the asymmetry of changes in utility that result from positive and negative changes in wealth or consumption of equal magnitude with a concave utility function, a risk-averse individual will always be willing to give up wealth in the good state of the world, i.e. reduce current consumption, and pay insurance against realization of the bad state in which wealth is low and marginal utility is high. In this context, the apparent question is what exactly causes the risk-neutral expected variance to frequently exceed the realized variance. A possible explanation is a phenomenon called asymmetric volatility (e.g. Wu (2001)) that describes a negative correlation between market returns and return variance. Due to this negative correlation, variance swaps naturally offer insurance against substantial market declines, which may explain why investors are willing to pay a premium for such instruments, i.e. accept a variance strike that is too high relative to the variance one would expect to realize under the physical measure.

Moreover, a randomly changing return variance also implies that an investor's final wealth is not only risky but also ambiguous[1] or uncertain and there is empirical evidence of ambiguity aversion, i.e. that people prefer to act on known rather than on unknown probabilities (Ellsberg (1961)[2]).

An interesting and intuitive explanation how model uncertainty, which is closely linked to ambiguity aversion, can also help to explain risk-neutral expected variances that frequently exceed realized variances is offered by Drechsler (2013). He develops a model in which the representative investor has a reference model about the evolution of certain economic fundamentals but is not confident whether this model is indeed correct. To account for this situation and to derive decisions that are robust to model uncertainty, the investor considers alternative models that are statistically difficult to distinguish from the reference model and evaluates his decisions under the model which represents the "worst case". The investor is particularly concerned that the reference model underestimates the intensity and magnitude of potential jumps. Thus, the risk-neutral probabilities that are determined under this "worst case" model are tilted toward states of the world in which wealth is low and marginal utility is high, in particular to states in which large negative jumps in the expected growth rate of cash flows occur. As such important shocks to the economic state affect asset prices and thus return variance, this mechanism directly translates into a higher risk-neutral expected variance. Because, irrespective of the sign, the realized return variance is positively affected by the occurrence of jumps, a variance swap obviously offers insurance against such adverse movements. Moreover, an implication of this model is that the

[1] Risk refers to a situation in which possible outcomes are uncertain but the distribution of outcomes is known. In contrast, ambiguity refers to a situation in which the outcome itself as well as the distribution of outcomes is not known (e.g Anderson et al. (2009)).

[2] Ellsberg (1961) describes a situation in which people can choose a ball to be drawn from two different urns that contain red and black balls. A certain prize is received if a red ball is drawn and a smaller prize if a black ball is drawn. Urn I contains exactly 100 balls but nothing is known about the relative proportion of red and black balls. Urn II contains exactly fifty black and fifty red balls. Once asked to bet on the outcome that a red (black) ball will be drawn, the majority of people prefer Urn II (Urn I) over Urn I (Urn II).

magnitude of the variance risk premium depends on the degree of investors' uncertainty. A circumstance that gives support to this model is that, once calibrated, it is able to generate the variance risk premium with a comparatively low level of risk aversion and to capture additional features of asset returns.[3] This is mainly a result of the explicit consideration of model uncertainty. In a related model by Drechsler and Yaron (2011) where this model uncertainty is not considered and the variance risk premium only arises from shocks to long-run consumption growth, larger jumps and a higher risk aversion are necessary to produce results that match important properties of the data. What both models (and many others) have in common, however, is that jumps are typically needed to produce variance risk premia that are close to the data estimates.

4. Methodology and data

4.1. Theoretical basis for extracting the risk-neutral expected variance

In this section, the methodology that is used to extract the risk-neutral expected variance from a set of traded options is outlined. The derivation of the risk-neutral expected integrated variance goes back to the work of Demeterfi et al. (1999), Britten-Jones and Neuberger (2000), Carr and Madan (1998) and others. Carr and Madan (1998) show that the integrated variance of the futures price process can be replicated through a static portfolio of European options across a continuum of strike prices and dynamic hedging in the underlying futures contract if continuous trading is possible, interest rates are constant and the underlying futures price is a continuous semi-martingale. Their solution builds on the work of Neuberger (1994) who finds that the payoff from delta-hedging a contract that pays the log of the spot price at maturity equals the difference between the realized variance and the variance assumed for hedging. Moreover, this log contract can be replicated through a static position in options (Breeden and Litzenberger (1978)).

In particular Carr and Madan (1998) derive the following expression for the time-0 conditional expectation of variance of the futures price process over the time horizon [0,T] under the risk-neutral probability measure $\mathbb{Q}$ where F_t denotes the time t price of the futures contract that expires at time T and $RV_{t,T}$ denotes the annualized realized variance over the time horizon [t,T].

$$E_0^{\mathbb{Q}}[RV_{0,T}] = E^{\mathbb{Q}}[\frac{2}{T}(log(\frac{F_0}{F_T}) + \frac{F_T}{F_0} - 1) - \frac{2}{T}\int_0^T (\frac{1}{F_0} - \frac{1}{F_t})dF_t] \quad (1)$$

The second term on the right hand side of equation (1) represents the payoff of a continuously rebalanced position in the underlying futures contract. The first term represents a log contract that pays $f(F_T)$ at maturity. The final payoff of this log contract can be expressed as the final payoff of a static portfolio of European out-of-the-money (OTM) put and call options on the underlying futures contract[4] across a continuum of strike prices K with expiration at time T where each option is inversely weighted by the square of its strike price K. Thus, the expression shown in equation (2) can be derived.

$$E_0^{\mathbb{Q}}[RV_{0,T}] = \frac{2}{T}E^{\mathbb{Q}}[\int_0^{F_0}\frac{1}{K^2}(K-F_T)^+dK + \int_{F_0}^{\infty}\frac{1}{K^2}(F_T-K)^+dK - \int_0^T(\frac{1}{F_0}-\frac{1}{F_t})dF_t] \quad (2)$$

Note that the terminal option payoff in equation (2) is the payoff on European options on the futures contract. However, since the options and the futures contract expire at time T and the futures price has to equal the spot price at maturity to prevent arbitrage, European options on the futures and options on the spot are effectively equivalent. Moreover, in a risk-neutral world, the drift of the futures price is zero when the money market account is taken as numeraire implying that the futures price is a martingale (Carr and Madan (1998)). Thus, the expected value from continuous trading in the futures contract is zero so that the last term in the square brackets can be omitted.

Since the current price of any asset should equal the present value of future expected cash flows, the final payoff of the option portfolio can be rewritten as the terminal value of the current price of the option portfolio. Thus, the time – t conditional expectation of variance over a time horizon $[t,T]$ under the risk-neutral measure can be stated as

$$E_t^{\mathbb{Q}}[RV_{t,T}] = e^{r_{t,T}(T-t)}\frac{2}{T-t}[\int_0^{F_0}\frac{1}{K^2}P_t(K,T)dK + \int_{F_0}^{\infty}\frac{1}{K^2}C_t(K,T)dK], \quad (3)$$

where $r_{t,T}$ is the time-t risk-free rate over the time horizon [t,T] and $P_t(K;T)$ and $C_t(K,T)$ are the time-t prices of European put and call options on the spot with strike price K that expire at time T. Equation (3) also denotes the variance

[3] In particular, the model is able to generate the high equity risk premium, low risk-free rate, excess volatility of equity returns relative to fundamentals, a substantial variance risk premium that has predictive power for equity returns and a realistic implied volatility surface that captures the implied volatility skew for different maturities.

[4] Carr and Madan (1998) formally show that any twice differentiable payoff function, $f(F_T)$, of the terminal futures price can be re-written as $f(F_T) = f(\kappa) + f'(\kappa)[(F_T-\kappa)^+ - (\kappa-F_T)^+] + \int_0^{\kappa} f''(K)(K-F_T)^+dK + \int_{\kappa}^{\infty} f''(K)(F_T-K)^+dK$. Setting the arbitrary parameter κ equal to F_0 allows to derive the expression shown in equation (2).

strike in a variance swap contract that is initiated at time t and expires at time T.

Equation (3) measures the risk-neutral expected variance exactly when the price of the underlying evolves according to a purely continuous price process without jumps. However, Jiang and Tian (2005) show that this expression also yields an accurate estimate under more realistic assumptions when the asset price process is allowed to exhibit jumps and the effect of implied jumps is included in the above measure so that an estimate of the total quadratic variation can be obtained. Further, Carr and Wu (2009) show in a simulation study that the jump-induced error is typically very small. In general, the approximation error due to jumps, i.e. the difference between the "true" risk-neutral expected variance and the estimate obtained from equation (3) will be positive when negative jumps dominate and can become significant when jumps significantly contribute to overall volatility, presumably in times of stress (Du and Kapadia (2011)).

Moreover note that equation (3) theoretically captures the risk-neutral integrated variance of the futures price process. However, under the assumption of deterministic interest rates and dividend yields, the spot and futures price should have the same quadratic variation. Moreover, since interest rates are at historically low levels for a substantial period covered by the dataset and the time horizon over which the risk-neutral expected variance is approximated is relatively short (22 trading days), a violation of these assumptions would probably have no significant effect on the results. Following the convention to determine the payoff on a variance swap that is described by Aït-Sahalia et al. (2015b), the annualized realized variance over the 22 trading day horizon $[t,t+22]$, where t now denotes a specific trading day, is defined as the annualized sum of daily squared log-returns shown in equation (4) based on a day-count-convention of 255 business days per year. The spot price used in equation (4) is adjusted for stock splits and spin-offs.

$$RV_{t,t+22} = \frac{255}{22}\sum_{i=1}^{22}(\log(\frac{S_{t+i}}{S_{t+i-1}}))^2 \tag{4}$$

As noted by Bollerslev et al. (2011), the use of model-free realized volatilities, computed by summing squared returns from high-frequency data, generally allows a more accurate ex post observation of historical volatility than the expression used in equation (4) and would therefore naturally lend itself for an accurate measurement of the variance risk premium as defined below. However, in an earlier version of their study, Bollerslev et al. (2008) find, based on the model of Heston (1993), that the mean bias in the volatility risk premium using model-free implied volatility and realized volatility estimated from daily returns was only about 1.05% of the theoretical volatility premium for a sample size of 600.[5] I consider this magnitude of bias sufficiently small to justify the use of equation (4) to measure the realized variance over the relevant horizon.

[5] This can be seen by comparing the figures in Table 1 of their paper. The mean bias in the volatility premium is about 0.0021 compared to a theoretical premium of -0.20.

4.2. The variance risk premium

Following Carr and Wu (2009), the variance risk premium is defined as the difference between the actually realized variance over the time horizon $[t,T]$ and the time-t expected variance under the risk-neutral measure over that same time period, i.e. $VRP_{t,T} = RV_{t,T} - E_t^{\mathbb{Q}}[RV_{t,T}]$. In this sense, $VRP_{t,T} \cdot 100$ is the payoff in monetary units to an investor who holds a long position in a variance swap contract with notional 100 that is initiated at time t and expires at time T, whereas $RVRP_{t,T} = \frac{RV_{t,T}}{E_t^{\mathbb{Q}}[RV_{t,T}]} - 1$ is the excess return to the investor when the variance swap rate is thought to be the initial investment. $LVRP_{t,T} = \log(RV_{t,T}/E_t^{\mathbb{Q}}[RV_{t,T}])$ can therefore be regarded as the continuously compounded excess return on this variance swap contract.

When no confusion arises, I drop subscripts in the text. Specifically, I use VRP, RVRP and LVRP to refer to $VRP_{t,T}$, $RVRP_{t,T}$ and $LVRP_{t,T}$ in the text.

4.3. Data

The applied dataset includes daily option price information on three US and five European stock indices as well as 29 single stocks from the US. A detailed overview of the included indices and stocks along with further information is shown in Table 1 below.

Daily settlement as well as bid and ask prices for options on European indices as well as the corresponding index series and all overnight indexed swap (OIS) rates are retrieved from Thomson Reuters Datastream. Bid and ask prices of options on US indices and stocks, the corresponding prices of the underlyings and the London Interbank Offered Rate (LIBOR) curve are retrieved from OptionMetrics. For the European indices, OIS rates are used as proxy for the risk-free rate. For the US underlyings, LIBOR rates are used as proxies for the risk-free rate prior to 2007 whereas from the year 2007 on, OIS rates are used as a proxy. The reason for this change is that derivative dealers generally used LIBOR as a proxy for the risk-free rate prior to 2007 but have switched to OIS rates for collateralized transactions in later years as a consequence of substantial rises in LIBOR rates during the financial crisis (Hull and White (2015)). Since the data used here stem from options traded on organized exchanges with central clearing authorities and collateral requirements, the application of these rates appears to be a reasonable choice.

If available, option prices for subsequent analyses are defined as the average of bid and ask prices. Otherwise, the daily settlement price is used. To obtain the final data sample, several exclusionary criteria are applied to the initial set of options. First, if applicable bid prices are required to be strictly positive and bid-ask-spreads to be equal or greater than zero. Second, options with less than seven days to maturity are excluded since prices of such options might be biased by liquidity and microstructure concerns

Table 1: Data description

Note: Entries list the name of included stocks and indices as well as their tickers which are sometimes used in other tables to reference the respective underlying. Start and End denote the first and last day in the sample on which a valid estimate of the variance risk premium could be obtained. NOptions denotes the number of option quotes available after filters are applied. NVSwap denotes the number of days on which a valid estimate for the VRP could be obtained. NKLow and NKHigh denotes the average daily number of options with expiration date T_L and T_H, respectively, from which implied volatilities could be derived and interpolated. TimeGap denotes the average distance in days between T_L and T_H over which the risk-neutral expected variance is interpolated. A description of T_L and T_H is given in section 4.4.

Underlying	Ticker	Start	End	NOptions	NVSwap	NKLow	NKHigh	Time Gap
AEX Index	AEX	06/06/2011	29/07/2015	54692	736	22.66	21.92	19.91
CAC40 Index	CAC	04/01/2010	29/07/2015	66668	967	23.95	18.84	19.93
DAX Index	DAX	18/04/2006	29/07/2015	257088	1660	44.43	45.23	19.73
Dow Jones Industrial Average	DJX	31/12/2002	30/07/2014	95091	2061	15.40	13.93	19.58
Euro Stoxx 50	ESX	03/01/2011	29/07/2015	114076	813	40.54	40.96	19.69
NASDAQ100	NDX	04/01/1996	30/07/2014	304070	3278	31.26	25.01	18.60
SMI	SMI	06/06/2011	29/07/2015	48250	717	22.97	18.99	19.49
S&P500 Index	SPX	04/01/1996	30/07/2014	649410	3598	47.88	42.58	16.43
Alcoa	AA	19/08/1997	30/07/2014	35747	803	5.13	6.56	18.46
Altria (Philip Morris)	MO	05/01/1996	30/07/2014	55143	2163	5.29	6.42	18.35
Amazon	AMZN	15/12/1997	30/07/2014	142691	2315	13.80	16.42	17.28
American Express	AXP	26/01/1996	30/07/2014	65397	1949	7.77	8.82	17.93
Amgen	AMGN	17/01/1996	30/07/2014	68730	2328	6.93	7.97	17.98
Analog Devices	ADI	20/08/1996	20/05/2014	29604	685	5.74	5.95	18.05
Apple	AAPL	04/01/1996	30/07/2014	272051	2392	23.22	28.25	16.81
Bank of America	BAC	17/01/1996	30/07/2014	51899	1467	5.22	6.40	15.60
Boeing	BA	27/02/1996	30/07/2014	77177	2032	7.77	8.83	17.24
Cisco	CSCO	04/01/1996	30/07/2014	63305	1932	6.23	7.84	17.76
Exxon Mobil	XOM	19/02/1997	30/07/2014	63227	1738	7.01	8.50	17.35
Facebook	FB	29/05/2012	30/07/2014	57679	501	20.35	21.38	8.22
General Electric	GE	17/01/1996	30/07/2014	60886	1860	5.63	7.18	17.14
Home Depot	HD	17/01/1996	30/07/2014	62245	1789	6.48	7.84	17.55
IBM	IBM	04/01/1996	30/07/2014	101647	2920	8.59	10.30	17.40
Johnson & Johnson	JNJ	26/01/1996	30/07/2014	47892	1238	5.92	6.88	16.92
McDonald's	MCD	29/02/1996	30/07/2014	56254	1406	7.01	8.13	16.49
Merck	MRK	22/02/1996	30/07/2014	62784	2038	6.38	7.60	17.96
Metlife	MET	13/02/2007	30/07/2014	37523	945	9.65	10.79	19.32
Microsoft	MSFT	04/01/1996	30/07/2014	87564	2441	7.78	9.91	18.13
Monsanto	MON	19/03/1996	30/07/2014	43459	1295	8.59	9.81	19.27
Nike	NKE	01/02/1996	30/07/2014	50464	1448	6.41	7.02	16.57
Pfizer	PFE	23/10/1996	30/07/2014	54150	1442	5.46	6.62	16.62
Procter & Gamble	PG	17/07/1996	30/07/2014	53938	1609	6.06	6.99	16.94
Starbucks	SBUX	18/09/1996	30/07/2014	53567	1530	7.06	8.41	18.92
Tesla	TSLA	21/07/2010	30/07/2014	64480	732	18.49	18.47	14.58
Valero	VLO	04/06/2001	30/07/2014	64166	1747	8.31	10.37	18.02
Verizon	VZ	25/03/1998	30/07/2014	59440	1399	7.12	8.69	16.65
WalMart	WMT	22/10/1997	30/07/2014	58038	1882	5.92	7.09	17.39

(Jiang and Tian (2005)). Third, index options with zero trading volume are excluded from the dataset because the prices of such options may not reflect true value (Jiang and Tian (2005)). Since most single stock options are less frequently traded than index options, a less strict criterion is applied to such options in that only options with zero trading volume and open interest smaller than 100 contracts are excluded.

Since the following applications require a dividend yield for the indices of which a reliable estimate is difficult to obtain, the option-implied dividend yield is obtained through a combined application of the put-call parity and spot-futures parity. In a first step, the time-t option-implied price, F_t^i, of the futures contract that expires at time T is inferred through put-call-parity (e.g. Hull (2009)) that is shown in equation (5).

$$F_t^i = K + e^{r_{t,T}(T-t)} \cdot (C_t(K,T) - P_t(K,T)) \tag{5}$$

In order to select the put and call option used in equation (5), the methodology applied by the Chicago Board Options Exchange for calculating its volatility index (CBOE (2015)) is used. For every maturity on a given day, the pair of options with the same strike price and maturity for which the absolute price difference is smallest is used in equation (5). The option-implied forward looking dividend yield at time t

over the time horizon $[t,T]$, $\theta_{t,T}$, is then extracted through the spot-futures parity (e.g. Hull (2009)) that is shown in equation (6).

$$F_t^i = S_t e^{(r_{t,T}-\theta_{t,T})(T-t)} \tag{6}$$

The outlined procedure ensures that the option pair used in the estimation of the implied dividend yield consists of at-the-money (ATM) options which are, especially for the indices considered here, typically relatively liquid. Thus, prices of such options and derived implied dividend yields can generally be expected to be reliable. Nevertheless, option-implied dividend yields are occasionally negative. Such negative values occur remarkably often during times of financial turmoil, especially during October 2008 and November 2011. Due to this, the negative yields most likely capture the effect of a discount rate used by market participants that exceeds the rate that is used to derive the implied dividend yield. Even though these negative dividend yields ensure that put-call parity is technically satisfied, it is obviously implausible to us them for further applications. Thus, the following adjustments are made to replace negative values: whenever possible, a replacement value for a negative forward looking dividend yield at a given maturity is obtained through linear interpolation or extrapolation of implied dividend yields at adjacent maturities on the same day. If a linear interpolation or extrapolation is not possible because the number of positive implied dividend yields at different maturities is smaller than two on a given day, the implied dividend yield for the same expiration date on the previous day is used as a replacement value. If none of these two adjustments results in a positive value or no value can be found, the implied dividend yield is set to zero. For those observations where the implied dividend yield is replaced, a new forward price corresponding to the adjusted implied dividend yield is computed.

For stocks a similar procedure is applied with the exception that the implied present value of dividends is derived instead of an implied dividend yield.

After the derivation of implied dividend yields and present values of dividends, a fourth exclusionary criterion is applied before implied volatilities are derived. As outlined by Aït-Sahalia and Lo (1998), in-the-money (ITM) options are traded relatively infrequently compared to ATM or OTM options. Therefore, prices of such options, and implied volatilities derived from these prices, tend to be unreliable. For this reason, ITM options are removed from the dataset.

For European options, implied volatilities are inferred based on the model of Black and Scholes (1973) (henceforth referred to as B-S model).[6] For American options, implied volatilities provided by OptionMetrics are used. For such options, OptionMetrics uses a binomial tree approach accounting for the effect of the early exercise premium.

Index options with implied volatilities greater than 80% and single stock options with implied volatilities greater than 100% are removed from the dataset. Such observations are probably outliers that would corrupt the implied volatility surface.

In addition to the options data, price information from the Center for Research in Security Prices, factor returns from Kenneth French's website and data from the Survey of Professional Forecasters, which are introduced in detail when applied, are used.

4.4. Implementation of the method to extract the risk-neutral expected variance

In order to obtain an estimate of the risk-neutral expected variance over the time horizon $[t,T]$, equation (3) is numerically evaluated using the trapezoidal method. A practical problem arises since equation (3) requires an infinite number of options across a continuum of strike prices whereas the number of traded options is finite. To cope with this situation, I generate 5000 artificial options that expire at the two expiration dates T_L and T_U closest below and above T over an equally-spaced range of strike prices of ±8 standard deviations from the current spot price. This standard deviation is estimated as the average of the implied volatilities of the two options that are closest to being at-the-money and expire at time T_L. In a simulation based on a model with stochastic volatility and jumps, Jiang and Tian (2005) show that truncation errors are generally negligible if the truncation points, i.e. the highest and lowest observed strike prices, are more than two standard deviations from the current forward price, F_0, and can be further reduced by applying the extrapolation scheme outlined below. They further show that discretization errors resulting from non-continuous strike prices are negligible when the gap between consecutive strike prices is smaller than or equal to 0.35 standard deviations. Thus, 5000 options over a range of strike prices between ±8 standard deviations from the current spot should help to alleviate truncation and discretization errors.

In order to generate the artificial option prices, implied volatilities are needed. As noted in section 4.3, implied volatilities for European options are obtained based on the model by Black and Scholes (1973) whereas implied volatilities provided by OptionMetrics are used for American options. For any given day t and the two expiration dates T_L and T_U, implied volatilities are then interpolated across moneyness, defined as $k \equiv \log(K/F_t^i)$, between the highest and lowest observed strike price using cubic splines. Jiang and Tian (2007) argue that the use of cubic splines leads to the convenient property that the implied volatility function is smooth over the range of observed strike prices, which is a direct implication of no-arbitrage constraints (e.g. Breeden and Litzenberger (1978)). For strike prices below the smallest and above the highest observed strike price, implied volatility is held constant at the value observed at these end points.[7]

[6] This is done with an adjusted version of Mark Whirdy's code "Fast Matrixwise Black-Scholes Implied Volatility" for Matlab that is available under the following address: http://www.mathworks.com/matlabcentral/fileexchange/41473-fast-matrixwise-black-scholes-implied-volatility.

[7] Jiang and Tian (2007) use a slightly different but similar linear extrapo-

The prices of artificial European options that expire at time T_L and T_U over the continuum of strike prices are obtained by inserting the interpolated implied volatilities into the B-S-formula. In the next step these option prices are used to numerically evaluate equation (3) and obtain an estimate of the risk-neutral expected variances over the two time horizons $[t, T_L]$ and $[t, T_U]$. An estimate of the risk-neutral expected variance over the desired horizon [t,T], can then be obtained through linear interpolation between the expected risk-neutral variances over the time horizons $[t, T_L]$ and $[t, T_U]$. Note that, in order to ensure that the interpolation and extrapolation scheme can work properly and derived estimates of the risk-neutral expected variance are reliable, this procedure is only applied on days on which at least four traded options are available. Moreover note that the B-S model is only used as a convenient way to derive implied volatilities from and translate implied volatilities into options prices. In particular, it is not assumed that this model is a true representation of reality.

A description of the average time over which the risk-neutral expected variance is interpolated along with the average number of options that expire at times T_L and T_H is shown in Table 1.

5. Empirical analysis

5.1. Realized variance risk premia

Figure 1 plots the daily time series of annualized risk-neutral expected and actually realized volatilities over the subsequent 22 trading days, the variance risk premium, defined as $VRP_{t,T}$, and the index level for the S&P 500 index between January 4th 1996 and July 30th 2014. From Figure 1 it is evident that the approximation of the expected variance captures the dynamic of the subsequently realized variance quite well and that substantial changes in the variance risk premium go hand in hand with significant fluctuations in the index level. This is also confirmed by the high correlation coefficient of 0.6437 between risk-neutral expected and subsequently realized variance for the S&P 500 index. At the same time, however, the wedge between the two variance measures is also apparent. For most of the time, the variance risk premium is negative and relatively stable but experiences substantial fluctuations when major market changes occur and reaches unprecedented levels during the financial crisis.

Table 2 shows summary statistics for the variance risk premium over a 22 trading day horizon in three different forms, defined either as the time series average of $VRP_{t,T}$ · 100, $LVRP_{t,T}$ or $RVRP_{t,T}$. The reason for introducing a logarithmic version of the variance risk premium is that the VRP and RVRP distributions exhibit substantial skewness and kurtosis, indicating that the two stem from a highly non-Gaussian distribution, while LVRPs generally appear more normal which might be beneficial for subsequent regression analyses.

Consistent with previous studies, the mean VRP is negative on all stock indices and statistically significantly different from zero at the 1%-level for the S&P 500, the AEX, and the Euro Stoxx 50 and significant at the 5%-level for the Dow Jones and NASDAQ 100. Only for the Swiss SMI and the DAX, the variance risk premium is not significant on any conservative level. The distributions of both, VRP and RVRP show substantial kurtosis and positive skewness and are reflective of high and positive returns which fatten the right tail of the distributions. In contrast, LVRPs exhibit a considerably lower skewness and kurtosis, which is due to the fact that especially the excessively high, positive returns during the recent financial crisis are alleviated through the logarithmic transformation. In this context, it is also noticeable that the standard deviations of the VRPs are excessive, especially for single stocks, which partly explains why the VRP is statistically different from zero only for comparatively few stocks. Indeed, even though VRP is negative in most cases, the null hypothesis that realized and expected risk-neutral variances do not differ on average cannot be rejected at the 5% significance level for 21 of the 29 stocks.

Mean LVRPs are negative for all underlyings and t-statistics[8] are generally higher. The mean LVRP is significantly different from zero at the 1%-level for all underlyings. However, as pointed out by Driessen et al. (2009), a necessary remark in this context is that due to the concavity of the logarithmic function, Jensen's inequality implies that mean log variance risk premia are negative even under the null hypothesis of equality between risk-neutral expected and subsequently realized variance. When raw variance swap returns ($RVRP_{t,T} = \frac{RV_{t,T}}{E^{\mathbb{Q}}[RV_{t,T}]} - 1$) are used instead, mean returns are generally smaller in absolute magnitude than for the LVRPs but still negative for 35 underlyings. However the mean raw return is statistically significantly different from zero at the 5%-level for only 14 underlyings.

Qualitatively, the results in the left panel of Table 2 are similar to those obtained by Carr and Wu (2009) who find significantly negative variance risk premia for stock indices but only comparatively few single stocks for which the premia are significant. Despite this tendency, the results are quite different from those of Driessen et al. (2009) who find a significantly negative variance risk premium for the S&P 100 index but no evidence for the presence of a generally negative variance risk premium in individual constituents. In-

lation scheme where the slope of the extrapolated segment is set equal to the slope of the interior segment at the endpoints. Even though this procedure better captures the observed skew in implied volatility surfaces for certain strike sections, it can lead to the implausible drawback that further-out-of-the-money call options have higher artificial prices than nearer-out-of-the-money call options, which occasionally happens here. Thus, the constant extrapolation scheme with zero slope is applied.

[8] Except for equation (9), all t-statistics and regressions, as well as standard deviations adjusted according to the method of Newey and West (1987) are obtained using Kevin Sheppard's MFE Toolbox for Matlab throughout this work. The toolbox is available under http://www.kevinsheppard.com/MFE_Toolbox.

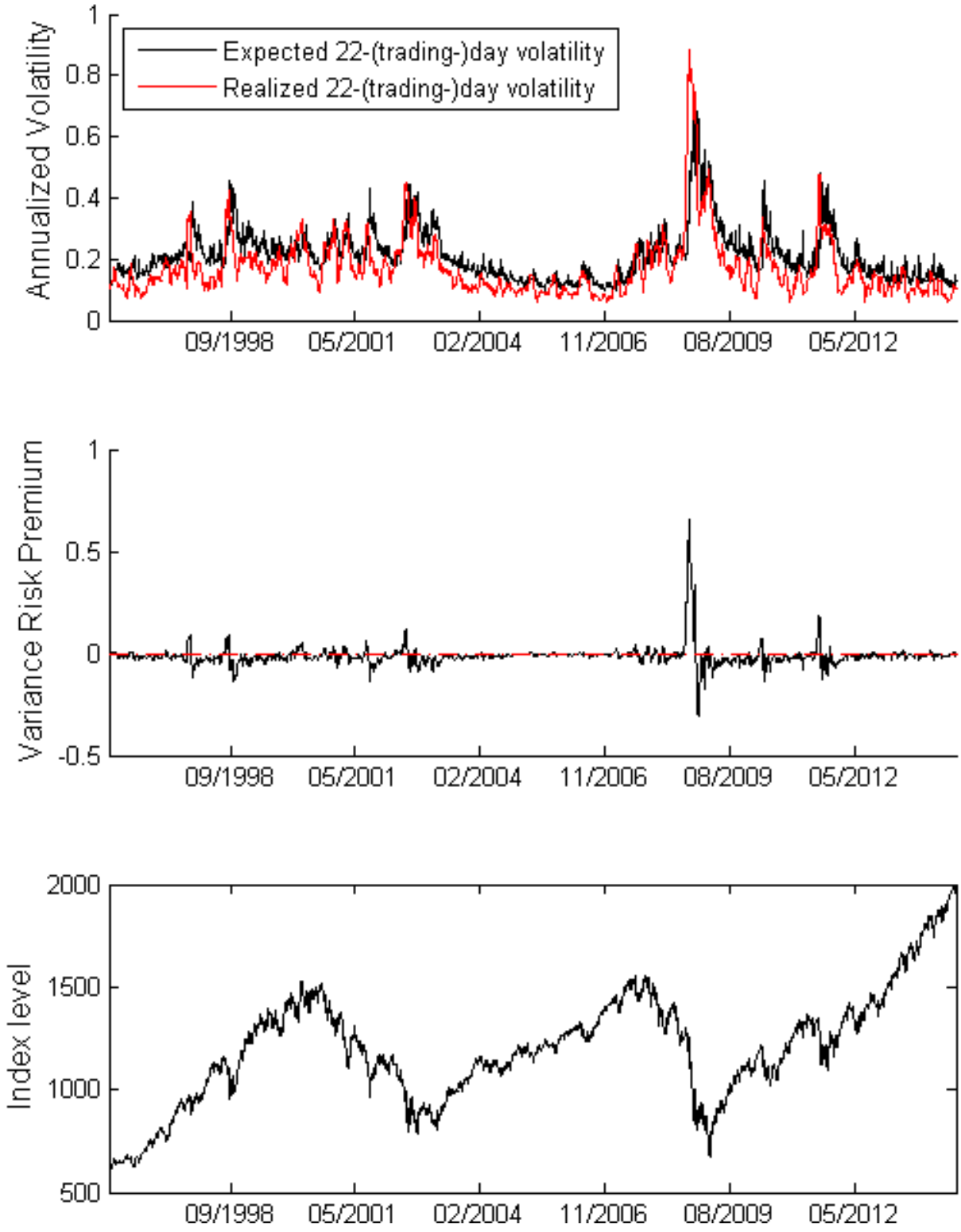

Figure 1: Time series of risk-neutral expected and realized volatility, variance risk premium and the S&P 500

Note: The first plot shows the time series of annualized risk-neutral expected and subsequently realized volatility over 22 trading days in the period from January 4th 1996 to July 30th 2014 for the S&P 500 Index. Both measures are annualized on a basis of 255 trading days in a year. The second plot shows the realized variance risk premium over 22 trading days, $VRP_{0,T}$, over the corresponding sample period. The red horizontal line indicates a VRP equal to zero. The third plot shows the index level of the S&P 500 index over the same sample period.

Table 2: Summary statistics of variance risk premia

Note: Entries refer to the time period from January 4th 1996 until July 29th 2015. RV and EV denote the realized and risk-neutral expected variance, respectively. Realized volatility is calculated from daily returns and annualized by $\sqrt{255/22}$. t-stat denotes the relevant t-statistics adjusted for serial correlation according to the method by Newey and West (1987) with a lag length of 22 days. Sample std., Skew, and Kurt denotes the sample standard deviation, skewness and kurtosis. Sharpe ratio is the annualized Sharpe ratio calculated as the annualized mean LVRP or RVRP divided by the corresponding sample standard deviation adjusted for serial dependence according to Newey and West (1987) with 22 lags and annualized by $\sqrt{255/22}$.

Descriptive statistics of variance risk premia

		VRP as $(RV - EV) \cdot 100$					LVRP as ln(RV/EV)						RVRP as RV/EV-1						
Ticker	Realized volatility	Mean	t-stat	Sample std.	Skew	Kurt	Mean	t-stat	Sample std.	Skew	Kurt	Sharpe Ratio	Mean	t-stat	Sample std.	Skew	Kurt	Sharpe Ratio	N
AEX	16.12%	-0.68	-3.01	2.18	1.22	8.51	-0.32	-4.61	0.59	0.43	3.27	0.58	-0.12	-1.70	0.65	2.56	11.20	0.11	736
CAC	19.88%	-0.62	-1.93	3.38	2.30	12.56	-0.27	-4.61	0.57	0.45	3.40	0.50	-0.09	-1.35	0.66	2.99	16.70	0.07	967
DAX	20.21%	-0.84	-1.65	5.78	4.02	37.84	-0.34	-6.46	0.61	0.50	3.85	0.54	-0.12	-1.91	0.74	3.51	19.47	0.06	1660
DJX	15.17%	-0.84	-1.99	4.95	4.92	51.04	-0.50	-10.72	0.58	0.74	5.04	0.80	-0.25	-4.63	0.67	4.46	28.23	0.14	2061
ESX	19.60%	-0.95	-3.02	2.88	1.30	7.73	-0.33	-5.05	0.55	0.27	2.96	0.60	-0.15	-2.60	0.53	1.91	7.70	0.18	813
NDX	26.05%	-1.17	-2.52	7.38	3.17	24.10	-0.33	-10.68	0.51	0.59	4.65	0.64	-0.17	-4.55	0.60	4.50	33.35	0.13	3278
SMI	14.05%	-0.24	-0.65	3.06	3.18	16.51	-0.33	-4.26	0.64	0.91	4.36	0.54	-0.08	-0.68	0.93	3.61	17.74	0.03	717
SPX	17.10%	-1.25	-3.65	5.15	5.72	60.03	-0.54	-15.05	0.57	0.79	5.07	0.85	-0.29	-6.61	0.67	5.61	46.08	0.14	3598
AA	40.18%	-0.37	-0.17	21.06	6.71	58.86	-0.22	-3.99	0.55	0.67	4.81	0.48	-0.04	-0.51	0.77	3.93	22.72	0.03	803
MO	25.68%	-1.16	-1.79	9.53	2.97	27.06	-0.37	-7.32	0.71	0.42	4.46	0.54	-0.07	-0.78	1.25	8.63	96.92	0.01	2163
AMZN	47.01%	1.65	1.01	24.60	2.74	14.05	-0.23	-5.31	0.62	0.45	3.91	0.38	-0.01	-0.23	0.86	4.11	27.72	0.01	2315
AXP	35.11%	1.13	0.90	15.47	4.26	26.86	-0.18	-4.34	0.57	0.64	4.44	0.33	0.00	0.01	0.80	3.53	18.79	0.00	1949
AMGN	32.06%	-1.54	-2.12	10.51	1.42	12.37	-0.30	-7.05	0.59	0.10	2.95	0.50	-0.12	-2.83	0.59	2.17	10.59	0.10	2328
ADI	51.22%	0.52	0.17	23.07	2.16	10.88	-0.19	-3.19	0.49	0.20	3.35	0.41	-0.06	-0.94	0.53	2.11	9.26	0.07	685
AAPL	39.31%	-1.31	-0.63	39.51	14.80	242.03	-0.31	-8.64	0.54	0.80	7.11	0.60	-0.11	-1.76	1.15	14.55	275.48	0.04	2392
BAC	39.00%	7.30	1.52	46.18	5.19	31.82	-0.19	-3.42	0.61	0.77	4.64	0.30	0.03	0.36	0.98	4.54	32.94	-0.01	1467
BA	28.94%	-1.35	-2.45	7.92	2.96	27.29	-0.26	-7.18	0.51	0.26	4.00	0.54	-0.12	-3.16	0.56	3.80	29.38	0.14	2032
CSCO	39.36%	-0.11	-0.10	14.34	1.92	10.54	-0.24	-5.38	0.62	0.29	3.22	0.42	-0.04	-0.73	0.72	2.55	11.85	0.02	1932
XOM	22.53%	-0.33	-0.33	12.02	9.05	98.71	-0.31	-6.73	0.57	0.72	5.26	0.55	-0.11	-1.45	0.90	7.55	84.13	0.04	1738
FB	43.97%	-2.21	-0.68	19.40	1.88	8.67	-0.25	-2.60	0.60	0.88	4.22	0.40	-0.03	-0.23	0.85	3.19	13.74	0.01	501
GE	30.84%	-0.82	-1.03	10.46	2.38	17.44	-0.23	-5.83	0.54	0.59	4.45	0.46	-0.06	-1.12	0.73	4.26	29.99	0.04	1860
HD	30.48%	-0.80	-1.01	12.18	7.80	93.72	-0.32	-7.60	0.55	0.49	4.77	0.61	-0.13	-2.54	0.78	7.69	87.12	0.09	1789
IBM	26.61%	-1.32	-3.14	6.87	1.22	12.72	-0.34	-9.08	0.58	0.22	3.21	0.57	-0.15	-4.08	0.58	2.49	12.76	0.13	2920
JNJ	19.07%	-1.26	-2.61	5.81	2.82	23.23	-0.43	-7.77	0.64	0.21	3.97	0.75	-0.19	-3.35	0.72	5.24	44.45	0.16	1238
MCD	18.85%	-1.29	-3.28	4.68	1.67	26.08	-0.41	-9.26	0.50	0.69	3.82	0.84	-0.24	-5.43	0.50	3.07	17.35	0.30	1406
MRK	26.06%	-0.81	-1.47	8.06	2.71	22.74	-0.29	-6.57	0.63	0.32	2.98	0.50	-0.08	-1.67	0.70	2.37	10.97	0.06	2038
MET	34.99%	1.02	0.35	30.18	10.83	140.21	-0.22	-4.11	0.53	1.35	9.33	0.46	-0.02	-0.16	1.28	9.28	104.98	0.00	945
MSFT	32.20%	-0.81	-1.20	9.94	3.09	20.24	-0.24	-6.53	0.55	0.34	3.39	0.45	-0.08	-1.86	0.62	2.71	14.56	0.06	2441
MON	32.56%	-1.04	-0.81	13.47	5.17	42.87	-0.29	-6.24	0.50	0.53	4.07	0.59	-0.14	-2.68	0.55	2.83	13.40	0.14	1295
NKE	29.08%	-0.99	-1.25	9.93	2.60	16.39	-0.31	-6.18	0.61	0.41	3.36	0.55	-0.10	-1.77	0.70	2.86	13.87	0.07	1448
PFE	27.28%	-1.32	-3.05	6.10	1.90	13.01	-0.30	-6.50	0.56	-0.01	3.78	0.58	-0.13	-2.96	0.56	2.56	13.24	0.16	1442
PG	21.38%	0.87	0.50	18.83	8.20	75.10	-0.39	-6.31	0.68	1.03	5.98	0.54	-0.06	-0.44	1.44	7.22	62.59	0.01	1609
SBUX	33.55%	-0.39	-0.43	12.89	4.39	37.57	-0.26	-5.59	0.56	0.51	3.28	0.49	-0.08	-1.67	0.65	2.77	14.51	0.07	1530
TSLA	54.35%	-5.68	-2.16	22.93	1.66	6.89	-0.32	-4.59	0.60	0.31	2.74	0.58	-0.12	-1.71	0.61	1.85	6.57	0.11	732
VLO	38.84%	0.12	0.08	17.15	5.42	44.52	-0.18	-4.17	0.52	0.48	3.68	0.34	-0.03	-0.62	0.62	2.68	13.15	0.02	1747
VZ	22.19%	-1.31	-2.11	6.85	3.76	40.91	-0.33	-8.23	0.49	0.41	3.72	0.75	-0.18	-4.31	0.50	3.38	22.50	0.25	1399
WMT	24.47%	-1.11	-2.31	6.02	1.09	12.29	-0.35	-8.19	0.54	0.32	3.30	0.64	-0.17	-4.20	0.54	2.56	12.94	0.18	1882

deed, the mean VRP is significantly negative at the 5%-level for only 7 of 129 (5%) stocks in their sample whereas this is the case for 9 of 29 stocks (31%) here. In this context, it is necessary to point out that the 29 stocks in this dataset have specifically been selected according to the number of price quotes provided by OptionMetrics and the number of valid price quotes following the criteria outlined above to allow the computation of a high number of estimates of the risk-neutral expected variance. Options that are comparatively illiquid and thus carry a substantial illiquidity premium should trade at lower prices, which would directly translate into lower implied volatilities. Applying the same procedure to derive the risk-neutral expected variance to such options should then naturally lead to lower expected risk-neutral variances and, holding realized variance constant, to less negative or even positive variance risk premia. Thus, the results here do not necessarily contradict the findings by Driessen et al. (2009) but could simply reflect different sample selection criteria and resulting differences in the sample composition.

When turning to the profitability of variance swap investments, the results show that the average continuously compounded return on shorting a 22-day variance swap on one of the major indices over the entire sample period is in the range of 27% in case of the CAC40 to 54% in case of the S&P 500. In this context, it is noticeable that the average returns on single stock variance swaps are lower than those on some indices, especially the US indices, but not substantially lower.

However, even though the variance premia are substantial, they are relatively low compared to the figures found by other studies. This circumstance is most obvious when one turns to the left panel of Table 2 where the payoff on a long variance swap for a notional of 100 currency units is shown. For instance, Carr and Wu (2009) find average payoffs of $2.5 and above for the S&P 500, the Dow Jones and NASDAQ. The lower payoffs over the entire sample might in part be due to the fact that especially during the financial crisis in 2008 and 2009 and its aftermath, realized variance has frequently and substantially exceeded expected variance.[9] This observation is also in line with the argument of Carr and Lee (2009) who point out that short positions in variance swaps led to significant and unprecedented losses especially during the final quarter of 2008. They further outline that market makers' difficulty to hedge their exposure to variance swaps, in particular for single names, led to a complete collapse of the single name variance swap market in 2009. For the sake of completeness, however, it is necessary to point out that the variance risk premia obtained here are generally smaller, in absolute terms, even over the same time period covered by Carr and Wu (2009).

[9] In addition, the smaller variance risk premiums found here compared to those found by Carr and Wu (2009) on the same data are due to the different interpolation schemes. While Carr and Wu (2009) use a linear interpolation scheme, the cubic spline interpolation applied here usually leads to lower implied volatilities and therefore lower option prices since the typical shape of implied volatilities across moneyness is convex. However, as shown below, the influence of the interpolation scheme is comparatively small.

In order to evaluate the profitability of variance swap investments with regard to their risk-return profile Table 2 also shows the annualized Sharpe Ratios. Since a variance swap contract is essentially a forward contract that does not tie up capital until maturity, the given variance swap returns directly represent excess returns and the Sharpe Ratios are simply computed as the annualized average return on the variance swap divided by the annualized return standard deviation which is again adjusted for serial correlation according to the method of Newey and West (1987). Considering the risk-return trade-off of shorting a 22-day variance swap, the high Sharpe Ratios for LVRPs of up to 0.85 for the S&P 500 index are relatively close to the estimates of Carr and Wu (2009) and suggest that variance swaps can be attractive investments. The premia that investors are obviously willing to pay for holding long variance swaps appear substantial compared to the risk to which the short side is exposed. However, when one considers the right panel in Table 2 for RVRPs, it is apparent that the high Sharpe Ratios for the LVRPs are mainly the result of the logarithmic transformation which not only alleviates the effect of large positive returns and thereby reduces the mean but also lowers the standard deviation. Both effects cause the Sharpe Ratios from shorting variance swaps to be substantially higher for LVRPs.

Of course, the extent to which the reported results are reflective of the profitability of actual variance swap investments crucially depends on how well the synthetic variance swap rates approximate the prices of actually traded instruments. However, at least for the S&P 500 index, Aït-Sahalia et al. (2015b) find that the methodology applied here leads to synthetic variance swaps rates for maturities of up to six months that are roughly in line with those of actually traded instruments.

As outlined before, the unprecedented levels of realized variance that manifested during the financial crisis also led to previously unseen returns on variance swaps. In order to evaluate the extent to which these returns affect the results reported in Table 2, Table A1 shows the same descriptive statistics for the time period from January 4th 1996 until December 31st 2007. For most underlyings the payoffs on a variance swap per 100 currency units notional is slightly higher in absolute terms than for the entire sample, whereas the continuously compounded returns as well as the raw returns are comparable and sometimes even lower in absolute terms than for the entire sample. The comparable mean returns over the two different samples can be explained by the fact that a long position in a variance swap earned substantial positive returns during the turmoil of 2008. However, after return variances had exploded, the variance swap rates increased accordingly and remained at their elevated levels even when realized variances reverted to lower levels so that average returns appear to be relatively unaffected over the entire sample period. This pattern is also apparent in the first plot in Figure 1. With regard to the relatively stable average variance swap returns, it is obvious that the Sharpe Ratios which are higher compared to those over the entire sample period for LVRPs as well as for RVRPs are predomi-

nantly driven by substantially lover standard deviations. Because the average returns are close to each other for the two samples, the sheer magnitude of variance swap returns that were realized during the financial crisis is disguised and it is not immediately apparent what could have led to the subsequent collapse of the single stock variance swap market mentioned by Carr and Lee (2009). In this context, it is helpful to consider the magnitude of variance swap returns only during this time period. The average maximum raw return on a long position in a 22-day single stock variance swap that was initiated during the final quarter of 2008 across all underlyings except the indices was a substantial 496.99%. For the S&P 500, the Dow Jones, and the NASDAQ, average 22-day returns on a long variance swap initiated during the last quarter of 2008 were 162%, 96%, and 107%. However, note that these figures are based on synthetic variance swap rates and especially when jumps significantly contribute to volatility, which was certainly the case at this time, the approximation error of the applied procedure to derive the risk-neutral expected variance can be high and the "true" risk-neutral quadratic variation is likely to be significantly underestimated (Du and Kapadia (2011)). Accordingly, actual variance swap rates would probably have been higher and realized variance swap returns lower. As a consequence, the stated figures can only serve as a rough indication of actual returns. Nevertheless, when considering the fact that market makers faced substantial problems in hedging their variance swap exposures at these times (Carr and Lee (2009)), the sheer magnitude of potential losses helps to understand why the single stock variance swap market could have broken down. All in all, it can be said that shorting variance swaps would have been a profitable strategy over the entire sample period despite the substantial losses during the financial crisis and its aftermath. However, when one considers raw returns, it is apparent that these losses have a deteriorating effect on Sharpe Ratios and these investments do not appear to be much more attractive than simple equity investments.

In order to ensure that the previously reported and following results are not substantially affected by the applied cubic spline interpolation scheme that is used to derive the risk-neutral expected variance and that has been chosen mostly independent from a guiding theory, a robustness test is performed. In order to do so, I recalculate the previously shown descriptive statistics based on a linear interpolation of implied volatilities in the range of traded strikes and apply the same extrapolation procedure as before. Results are shown in Table A2 in the Online-Appendix. As can be seen, the results under the two different interpolation procedures are almost identical. In general, the payoffs per 100 currency units of notional and the realized variance swap returns have a tendency to be slightly more negative when the linear interpolation scheme is used. This is due to the fact that the typical shape of implied volatilities across the applied moneyness measure is convex so that the cubic spline interpolation leads to lower interpolated implied volatilities on average. Therefore option prices are lower than under the linear interpolation scheme which then translates into lower estimates of the risk-neutral expected variance. However, the deviations are small and can reasonably be expected not to significantly alter the results.

5.2. Explaining variance risk premia

5.2.1. Exposure to overall market variance

From the previous results, it is obvious that the gap between risk-neutral expected and subsequently realized variance is persistent and statistically significant for indices as well as individual stocks. Investors are apparently willing to pay a significant premium for holding long positions in variance swaps. Since standard asset pricing theory postulates that only systematic risk factors can command such premia (e.g. Ross (1976)), it is instructive to investigate which systematic factors help to explain variance swap returns.

Payoffs on variance swaps naturally depend on the return variance of the underlying asset. However, only that portion of the asset's return variance that is systematic and cannot be diversified away in a portfolio of variance swaps can theoretically command a risk premium. Consequently, it is a reasonable starting point to test whether the covariation in an asset's return variance with that of the market portfolio helps to explain variance swap returns. This covariation is measured by a variance beta for each asset i as defined in equation (7). Because the distributions of return variances are heavily skewed, the logarithm of the return variance is used in equation (7), which allows to make the distributions more normal.

$$\beta_i^{var} = \frac{cov[\log(RV_i), log(RV_{market})]}{var[\log(RV_{market})]} \tag{7}$$

Initially, the variance beta for each stock or index is calculated over the entire horizon for which options data for the specific underlying are available. Over the entire common time period, the return variance over non-overlapping periods of five trading days is calculated for the underlying and the market portfolio. These figures are then used to determine the variance beta according to equation (7). Even though the variances are computed over a comparatively short period of only five trading days, the results are relatively robust to the use of alternative time horizons. The value-weighted portfolio of all stocks traded at the New York Stock Exchange (NYSE), American Exchange (AMEX) and NASDAQ is used as a proxy for the US market portfolio. Daily returns on this portfolio are obtained from the Center for Research in Security Prices (CRSP). For the three US indices and 29 individual stocks a regression of the mean LVRP over the entire horizon on the respective variance beta obtains the results shown in equation (8).

$$\begin{aligned} LVRP_i = &-0.2424 - 0.1044 \cdot \beta_i^{var} + e_i \quad R^2 = 4.94\% \\ &(-4.0327)\ (-0.8976) \end{aligned} \tag{8}$$

t-statistics are based on heteroskedasticity-robust standard errors and shown in parentheses. The goodness of fit is relatively poor with only 4.94%. The slope coefficient is negative but insignificant while the intercept is not substantially

smaller in absolute magnitude than the average risk premium and highly significant. Even though this result does not suggest any relation between market variance and the observed variance risk premium, it is misleading. The relation between the variance beta and the variance risk premium for the three US indices alone is virtually linear with an R^2 of 99.90% while the poor goodness of fit for equation (8) is mainly the result of an apparently non-existent relationship between exposure to market variance and realized variance risk premia for individual stocks. Of course, due to the limited number of only three index observations, this figure is far from meaningful.

In order to increase the number of observations, the regression approach is changed in the following way. For each of the three US indices, the mean logarithmic variance risk premium is calculated on a yearly basis for all calendar years between 1996 and 2014. The yearly mean logarithmic variance risk premium is defined as the average 22 trading-day variance swap return on all variance swaps whose initiation date falls into the respective calendar year. The corresponding variance beta for the specific year is calculated as outlined before using the returns that occurred in a particular calendar year.

Equation (9) shows the result for a pooled OLS regression with cluster-robust standard errors to control for the correlation between observations of the same index in different years.

$$\begin{array}{ll} LVRP_{it} = 0.1746 - 0.7018 \cdot \beta_{it}^{var} + e_{it} & R^2 = 16.66\% \\ \quad (1.79) \quad (-6.94) & N = 51 \end{array} \tag{9}$$

For the three US indices, there is a pronounced relation between the variance beta and the average logarithmic variance risk premium. The slope coefficient is negative and highly significant at the 1%-level whereas the intercept is not significant at the 5% level. These results are in line with the expectation that the average variance risk premium should be more negative for underlyings that have more exposure to overall market variance and suggest that exposure to overall market variance is indeed priced in US stock index options and variance swaps. However, the relatively low R^2 of only 16.66% shows that a significant proportion of overall variability in variance swap returns remains unexplained by the variance beta.

For a similar analysis with the five European stock indices, the Stoxx Europe 600 index is used as a common proxy for a European market portfolio. When the relation between the variance risk premium and the variance beta is examined over the entire common sample, the intercept is negative and highly significant whereas the slope coefficient is positive but insignificant. The positive slope coefficient is counterintuitive since it implies that investors are willing to pay a premium for additional exposure to systematic variance risk, a relation that is not supported by the assumption of risk-averse decision makers who dislike uncertainty. However, due to the limited number of only five observations this result should be considered with the appropriate degree of skepticism. In the corresponding regression with annual observations, the slope coefficient remains positive and insignificant.

In contrast, for the individual stocks, there is no discernible relation between exposure to overall market variance, as measured by the variance beta, and the variance risk premium. Estimating regression (8) and (9) for the stocks only leads to an insignificant slope coefficient and an R^2 indistinguishable from zero but a negative and highly significant intercept. Consequently, there is no evidence that exposure to systematic market variance is priced in variance swaps on individual stocks.

A possible explanation for this observation might be that the observed negative variance risk premium in options on single stocks does not, or only to a small extent, represent compensation for exposure to changes in market variance. There is an increasing amount of literature that documents a common factor structure in the volatility of idiosyncratic returns of stocks, i.e. the return component that cannot be explained by commonly applied factor models (e.g. Herskovic et al. (2014)), and increasing evidence that the risk of common changes in idiosyncratic volatility may indeed be priced (e.g. Gourier (2015), Cao and Han (2013)[10]). While standard-asset pricing theory would predict that idiosyncratic risk can be diversified away and should therefore not carry a risk premium, advocates of priced idiosyncratic risk argue that there might be structural issues, such as non-traded assets (Herskovic et al. (2014)), that cause market participants to hold undiversified portfolios. In the context of variance swaps, Schürhoff and Ziegler (2011) find a positive price of idiosyncratic variance risk whereas Gourier (2015) finds that idiosyncratic variance risk carries a negative price. In particular, Gourier (2015) finds that the idiosyncratic component of variance risk explains on average about 80% of the total variance risk premium. As a consequence, the apparently missing relation between exposure to market variance and the variance risk premia in stock options could possibly be explained if the observed premia entirely or in part represented compensation for changes in idiosyncratic variance. In order to test this hypothesis, I follow Schürhoff and Ziegler (2011) and use the excess returns on synthetic variance swaps on the S&P 500 Index as proxy for a systematic variance risk factor or price of overall market variance risk. The excess returns from variance swaps on single stocks and US indices are then regressed on this proxy. Using the variance swap returns on the S&P 500 index as a proxy for a systematic variance factor appears justifiable since the previous regression results show that returns on variance swaps on the three US indices are significantly associated with exposure to overall market variance. Moreover, in a simple factor decomposition Schürhoff and Ziegler (2011) show that index variances are predominantly determined by factor variances and the effect of idiosyncratic returns on index variances is negligible. If a significant proportion of single stock variance swap

[10] Cao and Han (2013) examine the returns on delta-hedged options that are sorted according to their idiosyncratic volatilities. They find that the average return from delta-hedging options on stocks with high idiosyncratic volatility is significantly lower than returns from hedging options on stocks with lower idiosyncratic volatility.

returns can then be explained by the systematic variance risk factor, these returns do probably not represent compensation for (common) idiosyncratic variance risk.

Table 3 shows the corresponding regression results. Except for Facebook, the slope coefficient on the systematic variance risk factor is positive and significant at the 1%-level for almost all regressions. R^2s frequently exceed 30% and achieve values of up to 93.20% and 70.06% for the two indices. This result clearly shows that variance swap returns on different underlyings have a pronounced tendency to move together. The systematic variance risk factor is obviously able to explain a substantial portion of the returns on single stock variance swaps. Even though the alpha is significant for several underlyings, including the two indices, it is not significantly different from zero at the 5%-level for 21 of 29 stocks. This finding supports the view that exposure to the systematic variance risk factor is mostly sufficient to explain variance swap returns and that the previously stated notion of priced idiosyncratic variance risk appears to play only a minor role for explaining single stock variance swap returns, at least for the sample at hand.

5.2.2. Commonly used risk factors

Sources of variance fluctuations and their implications for the pricing of variance risk

The results in the previous section suggest that returns on index as well as on single stock variance swaps have a strong tendency to evolve together and are captured well by the systematic variance risk factor. In this subsection, the objective is to further examine which risk factors precisely drive variance swap returns and to assess whether these returns can be explained by commonly used risk factors or whether exposure to changes in market variance indeed appears to command a separate risk premium. A prerequisite for variance risk to be priced as an independent risk factor is that return variance can evolve independently and affect aggregate consumption. Thus, it is important to consider the possible sources of changes in return variance and their implications for the pricing of variance risk.

Most reduced form option pricing models embed a negative correlation between equity returns and volatility (e.g. Cox (1996), Heston (1993)). Indeed, the presence of this negative correlation, often referred to as asymmetric volatility, is empirically well documented (e.g. Glosten et al. (1993)) and most apparent during significant market downturns that frequently come along with substantial increases in market volatility (Wu (2001)).

The phenomenon of asymmetric volatility is often explained by either the so-called leverage effect or the volatility feedback effect. According to the leverage effect hypothesis, a substantial stock price decline causes the market value of equity to fall more rapidly than the market value of a fixed amount of debt, thus increasing the debt-to-equity ratio. As a consequence the stock's risk rises, which in turn increases volatility (Black (1976)). Since the payoff of a variance swap is positively related to the level of realized variance, it obviously offers insurance against increases in market or stock volatility and, given the empirically documented negative correlation between market volatility and returns, also insurance against substantial declines in stock prices. This provides a rationale for the negative average return on long variance swaps. The important implication of the leverage effect hypothesis, however, is that volatility does not evolve independently but rather varies as a result of fluctuations in the stock price. If the functional relationship is even deterministic as for example in the model proposed by Dupire (1994), this also has important ramifications in the context of option pricing. An instantaneous volatility that evolves according to a deterministic function of the price of the underlying does not necessarily cause the market to be incomplete and can still allow to hedge an option solely with the underlying (Dupire (1994)). In such a setting, return variance does not represent an independent risk factor that could command a risk premium and the arguments outlined in section 3 that link the variance risk premium to market participants' inability to perfectly hedge an option position would not apply. Consequently, it may well be the case that the returns on variance swap contracts solely result from their directional stock price exposure rather than from explicitly priced variance risk. If this is indeed the case, the classical capital asset pricing model should be able to fully explain variance swap returns. A regression of variance swap returns on the excess market return can therefore reveal whether the observed variance risk premium reflects compensation beyond the correlation between market and variance swap returns.

The volatility feedback effect offers a second potential explanation for the observed negative correlation between volatility and equity returns. In this case, however, the causality between negative equity returns and increasing volatility is reversed as compared to the leverage effect. Here, volatility varies over time, for instance as a result of changes in business risk (Carr and Wu (2011)). Because volatility can fluctuate independently, it can theoretically command a risk premium and is indeed assumed to be priced. If volatility then rises, the required return on equity increases and stock prices fall, thus causing the negative correlation (Wu (2001)).

A third channel through which market returns can negatively interact with return variance is the so-called self-exciting behavior of financial markets which describes the phenomenon that substantial negative financial events seem to increase the likelihood of similar events to follow (Carr and Wu (2011)). Conceptually, this third channel is related to time-varying jump intensity (Carr and Wu (2011)).

Intuitively, all three explanations imply a negative market beta for variance swaps. However, only if volatility can fluctuate independently, it can constitute an additional risk factor that justifies a risk premium.

Capital Asset Pricing Model

In order to test whether the capital asset pricing model can explain observed variance risk premia, variance swap returns are regressed on the market excess return. For US underlyings, daily overlapping regressions as well as non-

Table 3: Regression of logarithmic variance risk premium on return on systematic variance risk factor

Note: Entries report the OLS estimates and t-statistics (in parentheses) of regressions of the continuously compounded excess return on variance swaps over a horizon of 22 trading days on Var_market. Var_market denotes the systematic variance risk factor or market price of overall market variance risk that is proxied by the continuously compounded excess return over a horizon of 22 trading days on synthetic variance swaps on the S&P 500 Index. t-statistics are adjusted for serial correlation according to the method of Newey and West (1987) with a lag length of 22 days. N denotes the total number of observations.

Underlying	$alpha$		Var_{market}		R^2	N
Dow Jones Industrial	0.031	(2.268)	0.961	(53.952)	93.20%	2049
NASDAQ100	0.059	(2.537)	0.747	(24.823)	70.06%	3265
Alcoa	-0.039	(-0.669)	0.463	(6.107)	32.68%	776
Altria (Philip Morris)	-0.169	(-2.719)	0.363	(4.712)	8.23%	2103
Amazon	-0.023	(-0.507)	0.395	(7.385)	13.45%	2211
American Express	0.158	(4.571)	0.695	(16.063)	55.36%	1864
Amgen	-0.066	(-1.638)	0.483	(10.918)	23.31%	2223
Analog Devices	0.056	(0.895)	0.493	(7.536)	36.17%	662
Apple	-0.078	(-1.921)	0.446	(8.047)	22.60%	2282
Bank of America	0.113	(1.879)	0.642	(9.273)	39.68%	1423
Boeing	-0.009	(-0.264)	0.485	(10.315)	33.04%	1966
Cisco	0.043	(1.028)	0.539	(11.479)	27.75%	1850
Exxon Mobil	0.027	(0.684)	0.671	(12.879)	51.36%	1679
Facebook	-0.280	(-2.408)	-0.050	(-0.229)	0.15%	501
General Electric	0.077	(2.462)	0.670	(16.918)	53.68%	1800
Home Depot	0.008	(0.215)	0.614	(13.373)	44.67%	1732
IBM	-0.024	(-0.624)	0.588	(13.531)	33.50%	2770
Johnson & Johnson	-0.067	(-1.375)	0.699	(10.159)	42.43%	1203
McDonald's	-0.182	(-4.333)	0.452	(9.918)	32.39%	1358
Merck	0.008	(0.172)	0.550	(11.732)	25.80%	1963
Metlife	0.153	(2.193)	0.621	(6.990)	51.90%	914
Microsoft	0.015	(0.421)	0.491	(8.943)	28.20%	2318
Monsanto	-0.028	(-0.688)	0.513	(9.936)	42.92%	1240
Nike	0.002	(0.029)	0.548	(8.604)	27.84%	1403
Pfizer	0.039	(0.924)	0.606	(10.807)	34.19%	1398
Procter & Gamble	-0.057	(-0.820)	0.649	(8.760)	34.88%	1555
Starbucks	-0.022	(-0.416)	0.442	(7.149)	23.10%	1496
Tesla	-0.303	(-3.156)	0.024	(0.205)	0.05%	718
Valero	0.031	(0.582)	0.400	(5.760)	20.96%	1674
Verizon	-0.083	(-1.857)	0.441	(7.363)	30.08%	1352
WalMart	-0.081	(-2.157)	0.537	(12.801)	35.18%	1815

overlapping regressions of 22 trading day variance swap returns on the corresponding excess market returns are conducted with the same proxy for the market portfolio as outlined above. The excess market return is calculated as the 22 trading day return on the market portfolio minus the four-week Treasury-Bill rate. For the regressions in Panel A of Table 4, the sample covers only the time period from July 31st 2001 until August 31st 2014, which is due to the fact that the daily time series of four-week Treasury bill rates is available only from July 2001 on. The results reported for regressions in Panel B generally cover the entire sample period for the respective underlying.

For the five European indices, results refer to non-overlapping monthly regressions of 22 trading day variance swap returns on the monthly excess return on a proxy for the European market portfolio. This excess return is publicly available at a monthly frequency from Kenneth French's online data library.[11] The regressions are estimated by OLS and standard errors are adjusted for serial correlation according to the method by Newey and West (1987) with 22 lags for the US underlyings in Panel A and unadjusted for the non-overlapping regressions shown in Panel B.

As predicted, the beta is negative and significant at the 1% level for almost all underlyings. The negative beta is

[11]The internet address is: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html#International. The countries considered in the European market portfolio proxy include Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom. This information is also available at Kenneth French's website.

Table 4: CAPM regressions

Note: Entries in Panel A report the OLS-estimates and t-statistics (in parentheses) of overlapping regressions of 22-day continuously compounded variance swap returns on the corresponding market excess return which is proxied by the 22-day excess return on the value-weighted portfolio of all stocks traded at the NYSE, NASDAQ, and AMEX. t-statistics are computed based on standard errors that are adjusted according to the method of Newey and West (1987) with a lag length of 22 days. Entries in Panel B report the OLS-estimates and t-statistics (in parentheses) of non-overlapping regressions of 22-day continuously compounded variance swap returns on the corresponding market excess return. For the US underlyings, the excess market return is proxied by the monthly excess return on the value-weighted portfolio of all stocks traded at the NYSE, NASDAQ, and AMEX. For the European indices, the market excess return refers to the excess return on a proxy for a European market portfolio taken from Kenneth French's website. t-statistics are not adjusted for serial correlation. N denotes the total number of observations.

	Panel A:						**Panel B:**					
Underlying	alpha	t-stat	beta	t-stat	R^2	N	alpha	t-stat	beta	t-stat	R^2	N
S&P 500 Index	-0.499	-15.476	-7.001	-7.393	37.30%	2587	-0.482	-14.006	-7.162	-8.177	31.60%	213
Dow Jones Industrial	-0.444	-12.835	-7.140	-6.997	36.60%	2056	-0.419	-10.012	-8.817	-7.189	39.10%	134
NASDAQ100	-0.357	-12.621	-5.863	-6.955	33.40%	2387	-0.305	-9.574	-6.254	-7.873	29.60%	207
Alcoa	-0.264	-6.189	-6.475	-8.655	39.80%	694	-0.237	-3.610	-6.953	-4.299	35.90%	50
Altria (Philip Morris)	-0.416	-6.228	-3.558	-2.971	5.60%	1464	-0.369	-6.660	-2.761	-2.391	3.60%	155
Amazon	-0.273	-6.259	-3.563	-4.869	8.00%	1978	-0.180	-3.362	-2.785	-2.485	3.60%	154
American Express	-0.198	-5.019	-5.436	-6.680	25.00%	1470	-0.118	-2.465	-4.356	-4.278	14.50%	129
Amgen	-0.332	-7.230	-4.255	-6.073	14.00%	1827	-0.263	-5.776	-3.587	-3.729	9.50%	154
Analog Devices	-0.278	-5.224	-2.790	-2.945	9.80%	467	-0.111	-1.497	-2.498	-1.974	6.50%	46
Apple	-0.319	-8.883	-4.396	-5.748	17.40%	1901	-0.260	-5.552	-5.473	-4.433	17.00%	153
Bank of America	-0.202	-3.667	-6.416	-6.601	30.20%	1094	-0.108	-1.868	-6.120	-4.842	25.70%	89
Boeing	-0.268	-7.969	-4.090	-6.849	19.70%	1721	-0.199	-4.800	-4.341	-5.430	17.00%	134
Cisco	-0.281	-5.135	-5.036	-5.136	18.20%	1135	-0.211	-4.220	-5.856	-6.476	21.40%	127
Exxon Mobil	-0.285	-7.204	-6.239	-6.582	33.30%	1573	-0.287	-5.845	-6.365	-4.865	28.40%	114
Facebook	-0.158	-1.728	-5.764	-2.180	6.00%	498	-0.265	-3.430	-1.911	-0.367	0.80%	24
General Electric	-0.251	-5.762	-5.241	-7.508	25.20%	1248	-0.205	-4.339	-3.801	-3.880	13.10%	119
Home Depot	-0.354	-9.578	-4.321	-7.794	19.60%	1429	-0.224	-4.232	-4.796	-5.474	16.80%	114
IBM	-0.394	-10.076	-4.495	-6.649	17.30%	1980	-0.316	-8.291	-4.685	-6.646	15.80%	190
Johnson & Johnson	-0.410	-7.405	-6.131	-5.423	26.60%	959	-0.442	-6.620	-3.496	-2.588	8.40%	82
McDonald's	-0.403	-10.156	-3.947	-4.251	19.40%	1291	-0.376	-7.209	-4.276	-3.254	16.90%	92
Merck	-0.321	-6.139	-4.241	-4.705	10.90%	1458	-0.255	-4.473	-3.510	-3.217	6.60%	131
Metlife	-0.160	-3.045	-5.765	-4.044	25.10%	943	-0.109	-1.426	-7.518	-3.034	29.30%	64
Microsoft	-0.260	-6.102	-4.088	-4.570	17.40%	1508	-0.200	-4.860	-3.910	-4.658	13.00%	160
Monsanto	-0.264	-7.202	-4.361	-5.492	23.90%	1284	-0.216	-4.409	-4.971	-4.186	21.80%	90
Nike	-0.293	-5.942	-5.222	-4.880	21.00%	1149	-0.226	-3.380	-4.502	-3.076	12.40%	87
Pfizer	-0.328	-5.584	-5.091	-5.498	18.90%	912	-0.237	-4.311	-2.937	-2.927	7.80%	94
Procter & Gamble	-0.463	-9.190	-4.592	-4.744	20.40%	1302	-0.363	-5.483	-4.735	-3.354	13.10%	97
Starbucks	-0.269	-6.179	-4.384	-5.491	15.40%	1287	-0.240	-4.497	-4.323	-3.854	13.30%	103
Tesla	-0.319	-4.777	-0.004	-0.003	0.00%	728	-0.393	-4.299	0.249	0.108	0.00%	45
Valero	-0.152	-3.944	-4.185	-5.471	15.90%	1741	-0.149	-2.938	-4.398	-3.504	11.50%	121
Verizon	-0.298	-8.135	-4.086	-5.130	20.50%	1352	-0.258	-4.686	-4.703	-3.822	20.40%	85
WalMart	-0.405	-11.140	-4.779	-8.185	26.50%	1462	-0.310	-6.541	-4.396	-5.186	16.10%	125
AEX Index	-	-	-	-	-	-	-0.258	-3.306	-9.058	-6.258	51.60%	26
CAC40 Index	-	-	-	-	-	-	-0.236	-3.778	-8.529	-7.756	61.10%	38
DAX Index	-	-	-	-	-	-	-0.311	-5.679	-7.061	-7.245	49.10%	63
Euro Stoxx 50	-	-	-	-	-	-	-0.267	-3.823	-8.460	-6.391	51.60%	30
SMI Index	-	-	-	-	-	-	-0.419	-3.613	-6.745	-3.341	23.40%	22

consistent with the empirically found negative correlation between return variance and equity returns. However, with a R^2 somewhere between 10% and 30% for the majority of regressions a substantial portion of the variation in variance swap returns remains unexplained. Furthermore, with only few exceptions, the alpha is negative and highly significant at the 1%-level, with the absolute magnitude of the alpha not significantly lower than the average variance risk premium reported in Table 2. Indeed, based on the average 22 trading day market excess return for the US market of approximately 0.89% over the sample period in Panel A, the market beta accounts on average for only approximately 7.01%[12] of the logarithmic variance risk premium across all US underlyings. A comparison of the two panels moreover shows that the results are qualitatively unaffected by the sample frequency.

For the five European indices, the situation is very similar. The market beta as well as the alpha is negative and highly significant leading to the same conclusion as for the US underlyings. What is striking, however, is that the R^2*s* for the European indices are mostly in the range of 50% and therefore substantially higher than for the US indices, which suggests that the selected market portfolio proxy is appropriate and captures the dynamic in European variance swap returns quite well.

All in all, the fact that the market beta cannot fully account for the realized variance swap returns represents an indication that the observed variance risk premium reflects compensation for other factors than the correlation between variance swap returns and market excess returns.

Fama-French three-factor model

Considering the fact that the CAPM has been found to be insufficient to explain stock returns (Fama and French (2004)), there is no apparent reason to believe the CAPM should perform much better in explaining variance swap returns. Thus, in order to evaluate whether these returns can be better explained when additional factors are considered, they are regressed on the excess market return and the two additional factors identified by Fama and French (1993), a firm size factor (SMB) and a book-to-market factor (HML). The excess return realizations for the market portfolio proxy and the SMB and HML factor portfolios, both for the US and Europe, are publicly available at Kenneth French's website at a monthly frequency.[13] The market portfolio proxies are the same as those used for the CAPM regressions. Since the excess returns are published on a monthly basis, the regression results in Table 5 refer to monthly non-overlapping regressions.

The magnitude of neither the alpha nor the market beta changes substantially compared to the values obtained in the CAPM regressions. Moreover, both parameters remain highly significant. Consistent with the findings of Carr and Wu (2009), the loading on the SMB factor is usually negative and often significant for the US underlyings, which suggests a negative correlation not only between return variance and market returns but also between return variance and returns on the SMB factor. Long variance swaps thus appear to offer insurance against a situation in which small stocks underperform relative to large stocks. The sign of the loadings on the HML factor is less persistent and the loading is mostly insignificant. For the European indices the situation is again very similar. Altogether, the persistently negative regression alphas suggest that variance swap returns reflect compensation beyond the swap's pure correlation with the market portfolio and that commonly used risk factors are not able to explain this additional compensation. The results therefore point towards additional risk factors that command a persistent and economically substantial risk premium.

5.2.3. Jump and variance risk

Construction of risk-factor-mimicking portfolios

Thus far, commonly used risk factors do not appear to be able to satisfactorily explain variance swap returns, which is indeed supportive of the notion that these returns reflect compensation for one or more additional priced risk factors. As a consequence, I construct alternative factor portfolios that allow to gain separate exposure to the risks that theory suggests to be priced in option returns and evaluate whether the returns on these risk factor portfolios help to explain variance swap returns.

The total realized quadratic variation is the result of continuous as well as of discontinuous price movements and the uncertainty associated with each of these kinds of movements may induce investors to command separate risk premia for them. Indeed, as outlined in section 3 and noted by Bollerslev and Todorov (2011), the commonly observed variance risk premium does not only reflect compensation for stochastic volatilities but also for time-varying jump intensities and fears for jump tail events. The latter can explain up to three quarters of the variance risk premium (Bollerslev and Todorov (2011)). Especially the fact that jumps are typically required in models to generate a variance risk premium whose magnitude is reconcilable with that observed in reality (e.g. Drechsler (2013)) or in option pricing models to fit certain characteristics of the cross-section of option prices (e.g. Pan (2002)) may point to the importance of jump risk when explaining the variance risk premium.

Thus, to examine the degree to which the returns on variance swaps represent compensation for presumed jump risk, i.e. risk associated with discontinuous price movements, and diffusive risk, I construct two factor portfolios from traded options that are intended to offer exposure to one of the two risk factors while being relatively unaffected by the other. For this purpose, I follow Cremers et al. (2015) who use the same factor portfolios to examine the influence of jump and volatility risk on the cross-section of stock returns.

On any given day t in the sample, I construct two delta-neutral straddles with different maturities. For the US un-

[12] For each underlyings, this contribution is calculated as $marketbeta \cdot 0.89\%/(\alpha + marketbeta \cdot 0.89\%)$.

[13] The internet address is: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Table 5: Fama-French Three-Factor regressions

Note: Entries report the OLS-estimates and t-statistics (in parentheses) of monthly, non-overlapping regressions of 22-day continuously compounded variance swap returns on the corresponding excess return on the market portfolio and the two additional factors, SMB and HML, identified by Fama and French (1993). For the US, the market excess return is proxied by the 22-day excess return on the value-weighted portfolio of all stocks traded at the NYSE, NASDAQ, and AMEX. For the European indices, the market excess return refers to the excess return on a proxy for a European market portfolio retrieved from Kenneth French's website. t-statistics are not adjusted for serial correlation. N denotes the total number of observations.

Underlying	alpha	t-stat	market	t-stat	SMB	t-stat	HML	t-stat	R^2	N
S&P 500 Index	-0.474	-13.818	-7.020	-7.642	-2.073	-1.585	-1.780	-1.578	33.20%	213
Dow Jones Industrial	-0.423	-10.521	-9.717	-7.252	1.164	0.636	3.798	1.840	41.10%	134
NASDAQ100	-0.299	-9.323	-5.848	-7.217	-2.552	-2.893	-0.832	-0.808	31.90%	207
Alcoa	-0.268	-4.114	-6.933	-4.305	-2.553	-1.530	4.212	2.364	44.10%	50
Altria (Philip Morris)	-0.345	-5.994	-2.871	-2.433	-2.698	-1.913	-3.409	-2.176	6.40%	155
Amazon	-0.174	-3.309	-2.160	-1.831	-2.967	-1.860	-4.557	-2.476	7.90%	154
American Express	-0.117	-2.460	-4.296	-4.222	-0.978	-0.715	-1.495	-1.060	15.30%	129
Amgen	-0.253	-5.641	-3.408	-3.362	-1.988	-1.240	-2.081	-1.285	11.40%	154
Analog Devices	-0.067	-0.904	-2.043	-1.629	-3.645	-1.987	-2.790	-1.696	15.00%	46
Apple	-0.261	-5.877	-5.282	-4.240	-0.680	-0.573	0.291	0.169	17.20%	153
Bank of America	-0.093	-1.697	-6.328	-5.769	-2.642	-2.138	-4.000	-2.744	32.30%	89
Boeing	-0.202	-4.903	-3.753	-4.321	-2.962	-1.744	-0.366	-0.271	18.90%	134
Cisco	-0.216	-4.439	-5.654	-6.032	-0.140	-0.105	1.332	1.081	22.10%	127
Exxon Mobil	-0.286	-5.783	-6.261	-4.543	-0.581	-0.334	-0.055	-0.030	28.50%	114
Facebook	-0.287	-3.932	-2.697	-0.561	2.281	0.443	6.983	1.823	4.90%	24
General Electric	-0.202	-4.350	-3.885	-4.134	-2.034	-1.553	-2.115	-1.875	16.00%	119
Home Depot	-0.220	-4.059	-4.465	-4.996	-1.242	-0.736	0.261	0.221	17.60%	114
IBM	-0.308	-8.118	-4.378	-6.029	-2.629	-2.231	-1.391	-1.148	18.30%	190
Johnson & Johnson	-0.421	-6.001	-3.530	-2.369	-2.622	-1.162	-2.520	-1.102	11.60%	82
McDonald's	-0.379	-7.339	-4.638	-2.927	1.083	0.677	1.632	0.601	17.90%	92
Merck	-0.257	-4.474	-3.399	-2.827	-0.123	-0.064	0.575	0.344	6.70%	131
Metlife	-0.107	-1.399	-8.161	-2.621	0.820	0.244	3.120	0.799	31.00%	64
Microsoft	-0.192	-4.874	-3.594	-4.154	-2.993	-2.803	-1.209	-0.988	16.50%	160
Monsanto	-0.216	-4.421	-4.154	-3.008	-2.919	-1.337	-0.847	-0.356	23.40%	90
Nike	-0.244	-3.734	-3.616	-2.493	-5.066	-2.219	-2.298	-1.161	16.50%	87
Pfizer	-0.230	-4.355	-2.867	-2.946	-2.519	-2.051	-2.429	-1.834	11.80%	94
Procter & Gamble	-0.357	-5.439	-4.066	-2.583	-4.183	-1.158	0.570	0.321	18.90%	97
Starbucks	-0.238	-4.362	-4.597	-4.080	0.325	0.249	-2.208	-1.465	15.30%	103
Tesla	-0.373	-3.771	-0.975	-0.344	4.394	0.802	-5.421	-0.874	4.80%	45
Valero	-0.150	-3.052	-5.179	-3.570	2.468	0.971	1.470	0.680	12.60%	121
Verizon	-0.251	-4.420	-4.725	-3.470	-1.505	-0.678	1.911	0.981	21.80%	85
WalMart	-0.304	-6.276	-4.369	-4.858	-0.952	-0.475	-0.996	-0.703	16.60%	125
AEX Index	-0.246	-4.060	-8.005	-4.824	0.083	2.233	-0.021	-0.540	58.60%	26
CAC40 Index	-0.237	-4.577	-8.238	-6.063	0.045	1.426	-0.004	-0.142	63.00%	38
DAX Index	-0.287	-5.337	-8.039	-7.546	0.038	1.160	0.033	0.997	50.70%	63
Euro Stoxx 50	-0.251	-4.033	-8.167	-5.347	0.056	1.529	0.000	-0.014	54.80%	30
SMI Index	-0.396	-4.317	-5.731	-2.400	0.091	1.628	-0.025	-0.451	31.40%	22

derlyings, these straddles are constructed from options on the S&P 500 index. The choice falls on this specific index because the corresponding options are highly liquid, have the highest number of valid price quotes of all underlyings and cover the longest time period in the sample. Moreover, since the S&P 500 represents the most diversified portfolio among the three US indices considered here, straddles constructed from options on it are most likely to capture exposure to purely systematic risks. In case of the European indices, separate straddle portfolios are constructed for each of the indices from options on the respective index. In general, the same criteria apply as those outlined in section 4.2 with the exception that in-the-money options are not excluded since the construction of straddles involves put and call options with identical strike and time to maturity that are close to being at-the-money.

Following Cremers et al. (2015), the straddle with the shorter maturity is constructed from the call and put options that are closest to being at-the-money and which expire in the calendar month following the month during which the straddle is constructed while the at-the-money options used to form the straddle with the longer maturity expire in the calendar month that follows. Each straddle return is constructed by solving the following problem

$$w^c_{t-1}\delta^c_{t-1} + (1 - w^c_{t-1})\delta^P_{t-1} = 0, \tag{10}$$

$$R_t^S = w_{t-1}^c R_t^C + (1 - w_{t-1}^c) R_t^P, \quad (11)$$

where w_t^c denotes the weight of the call in the straddle, R_t^C and R_t^P denote the one-day returns on the call and put, respectively, δ_t^c and δ_t^P denote the Black and Scholes (1973) deltas relative to the option price of the call and put and R_t^S is the weighted one-day return on the straddle, all at time t. In a second step, the JUMP factor which is intended to capture jump risk, and the VOL factor which is intended to capture the diffusive component of volatility risk, are constructed as a weighted combination of the previously created delta-neutral straddles with different maturities.

In order to construct the JUMP factor, a long position in one short-maturity straddle is combined with a short position in n long-maturity straddles such that the combined position is vega-neutral and thus insensitive to moderate changes in implied volatility. Analog to Fournier and Jacobs (2015), the one-day return on the JUMP factor, $JUMP_t$, is thus given by

$$JUMP_t = R_t^{S1} - \frac{Vega_{t-1}^{S1}}{Vega_{t-1}^{S2}} R_t^{S2}, \quad (12)$$

where R_t^{S1} and R_t^{S2} denote the one-day returns on the short- and long-maturity straddle, respectively and $Vega_t^{S1}$ and $Vega_t^{S2}$ denote the Black and Scholes (1973) vegas of the short- and long-maturity straddles relative to their market prices, all at time t. $JUMP_t$ can also be interpreted as the return on a portfolio that consists of a long position in the short-maturity straddle worth \$1 and a short position in the long-maturity straddle worth $Vega_{(t-1)}^{S1}/Vega_{(t-1)}^{S2}$ dollars. Since the vega of an option increases with time to maturity, the number of long-maturity straddles, n, is always lower than the number of short-maturity straddles held in the combination. At the same time, the gamma of an at-the-money option is higher when the time to expiration is lower. Thus, the combined position of the two delta-neutral straddles has positive gamma. In particular this positive gamma constitutes the theoretical exposure to jumps. Whereas the option delta postulates a linear relation between the option price and the price of the underlying, the option gamma measures the convexity in this relation. Due to this convexity, the established hedge is not perfect (Hull (2009)) and the corresponding hedging error increases with the magnitude of changes in the price of the underlying. Moreover, this hedging error always works in favor of a long position in the constructed portfolio since, holding everything else constant, a given increase in the price of the underlying causes a stronger price appreciation in the short-term straddle (that is held long) than in the long-term straddle (that is held short). Similarly, a given reduction in the price of the underlying causes a smaller price depreciation in the short-term straddle than in the long-term straddle. At the same time, however, the theta of an at-the-money option is usually more negative when the remaining time to expiration is shorter (Hull (2009)). Because the theta of the short-term straddle is more negative than that of the long-term straddle and the number of short-term straddles held is larger, the overall combination has a negative theta. Thus, holding everything else constant, the option portfolio will lose value from one day to the next. Consequently, the overall return on the jump factor will usually be negative and positive only when the hedging error overcompensates for the value reduction that results from the time decay. This will usually be the case when comparatively large movements in the price of the underlying, i.e. jumps, occur.

The VOL factor, in contrast, is specifically designed to be relatively unexposed to jump risk and to only capture the diffusive component of volatility risk. In order to achieve this goal, it is constructed to be delta-neutral and gamma-neutral but to have positive vega. Thus, the VOL factor consists of a long position in one long-term straddle and a short position in n short-term straddles so that the one-day return on the VOL factor, VOL_t, is given by

$$VOL_t = R_t^{S2} - \frac{Gamma_{t-1}^{S2}}{Gamma_{t-1}^{S1}} R_t^{S1}, \quad (13)$$

where $Gamma_t^{S1}$ and $Gamma_t^{S2}$ denote the time-t Black and Scholes (1973) gammas of the short- and long-maturity straddles relative to their prices.

Empirical characteristics of risk-factor-mimicking portfolios: Average returns

Table 6 shows descriptive statistics for the JUMP and VOL factor portfolios for both discrete and continuous returns. In case of the factor portfolios for the US, the JUMP factor earns substantial negative average returns for both forms, discrete and continuous, even though they are significantly negative only for the continuous form. Note, however, that the insignificance for the discrete form of the JUMP factor is due only to two extreme outliers of +83.68% and +45.93%. Excluding these returns leads to a t-statistic of -2.096.

The picture is similar for the VOL factor which also earns significantly negative average returns although the magnitude is somehow smaller than for the JUMP factor. The low correlation between the two factors of only -0.14 moreover suggests that they capture mostly unrelated return sources. Nevertheless, the negative correlation between the two factors is somewhat surprising since for example Todorov (2010) finds that price jumps are associated with corresponding spikes in stochastic volatility.

Figure 2 shows the cumulative returns on the JUMP and VOL factors over the period from January 5th 1996 to August 29th 2014 and further illustrates the economic magnitude of the negative average returns investors would have incurred if they had revolved the positions daily.

Even though long positions in both factors would have earned substantial positive returns around some major events such as the Asian crisis, or the bankruptcy of Lehman Brothers, the cumulative returns quickly converge to -100%. Considering the fact that both factor are market-neutral per construction, these results suggest that the JUMP and VOL factors indeed capture different factors that are both heavily priced by the market, presumably jump risk and volatility risk.

Table 6: Descriptive statistics for the risk factor mimicking portfolios JUMP and VOL

Note: JUMP denotes the daily return on a delta-neutral, vega-neutral and positive gamma calendar spread option strategy. VOL denotes the daily return on a delta-neutral, gamma-neutral and positive vega calendar spread option strategy. The additives (disc.) and (cont.) indicate the use of discrete and continuously compounded daily returns, respectively. t-statistics are shown in parentheses. N denotes the number of observations and SD denotes the standard deviation.

			Annualized Mean		Annualized SD	Annualized Sharpe Ratio	Daily Median	Skewness	Kurtosis	N
S&P 500	JUMP	(disc.)	-0.172	(-1.347)	0.544	-0.317	-0.005	5.157	95.821	4597
		(cont.)	-4.856	(-51.766)	0.398	-12.191	-0.016	-1.759	25.465	4597
	VOL	(disc.)	-0.138	(-2.332)	0.252	-0.549	-0.001	2.376	46.496	4597
		(cont.)	-0.122	(-2.011)	0.258	-0.474	-0.001	0.731	35.327	4597
AEX	JUMP	(disc.)	0.213	(0.877)	0.496	0.431	-0.004	3.322	39.396	1057
		(cont.)	-4.318	(-24.836)	0.354	-12.193	-0.014	0.583	32.682	1057
	VOL	(disc.)	0.185	(1.536)	0.245	0.754	-0.001	5.804	80.325	1057
		(cont.)	0.158	(1.418)	0.226	0.696	0.000	4.529	57.356	1057
CAC40	JUMP	(disc.)	0.591	(2.231)	0.478	1.236	-0.003	1.862	8.917	829
		(cont.)	-5.136	(-26.264)	0.353	-14.558	-0.017	-1.516	8.755	829
	VOL	(disc.)	-0.052	(-0.477)	0.196	-0.264	-0.001	0.564	6.292	829
		(cont.)	0.014	(0.135)	0.193	0.075	0.000	0.720	8.169	829
DAX	JUMP	(disc.)	0.016	(0.100)	0.480	0.033	-0.004	4.836	60.087	2306
		(cont.)	-4.576	(-37.536)	0.367	-12.480	-0.015	-0.789	112.567	2306
	VOL	(disc.)	0.038	(0.627)	0.182	0.208	0.000	2.111	37.169	2306
		(cont.)	0.102	(1.361)	0.225	0.453	0.000	11.306	361.090	2306
Eurostoxx	JUMP	(disc.)	0.229	(1.003)	0.485	0.471	-0.005	2.947	28.961	1153
		(cont.)	-4.497	(-27.107)	0.353	-12.742	-0.015	-0.347	33.886	1153
	VOL	(disc.)	-0.079	(-0.806)	0.210	-0.379	-0.001	-1.068	49.187	1153
		(cont.)	-0.041	(-0.446)	0.197	-0.210	-0.001	0.604	33.659	1153
SMI	JUMP	(disc.)	1.180	(2.277)	0.650	1.816	-0.003	3.712	30.644	400
		(cont.)	-5.868	(-15.198)	0.484	-12.120	-0.019	-3.473	34.358	400
	VOL	(disc.)	0.118	(0.572)	0.260	0.456	-0.001	0.464	8.787	400
		(cont.)	0.024	(0.115)	0.260	0.092	0.000	-1.157	19.515	400

Moreover, the negative sign of the average returns is consistent with models where investors dislike jumps or fear a deteriorating investment opportunity set due to increased variance and are therefore willing to accept lower or even negative average returns on assets or combinations of such assets that offer a hedge against jumps or increased variance (e.g. Drechsler (2013)).

For the five European indices, the picture is remarkably different. Discrete returns on the JUMP factor are on average positive for all indices even though insignificant for three of them. However, for the CAC40 and the SMI, the annualized returns are substantial and significant. In general, positive average returns on the JUMP and VOL factors are counterintuitive and contradict the idea that investors are willing to accept lower average returns on assets that offer insurance against undesirable states of the world. In this context, a possible bias that results from the factor portfolio construction methodology should be taken into account. This bias could particularly concern the risk factor returns for the CAC40 and SMI. As can be seen in Table 6, the two indices for which average factor portfolio returns are remarkably high (CAC40 and SMI) are also the two indices with the lowest number of valid days on which straddle returns could be computed. This situation occurs even though the raw data on the AEX and SMI cover exactly the same time horizon and the data on the CAC40 cover an even longer time period than on the Euro Stoxx 50. The reason for the different numbers of valid days is that the straddles that are used for the two factor portfolios can only be formed when the same ATM-options have positive trading volume on two consecutive trading days, which apparently is not always the case. If one now suspects that investors' trading in options is, at least in part, driven by the desire to insure against substantial negative price movements or rising volatility or otherwise take speculative positions, investors are probably more likely to trade immediately around the time of certain events, for instance when the underlying has incurred unexpected price fluctuations or volatility has unexpectedly risen. Indeed, Cao and Ou-Yang (2009) conclude that differences of opinion are more likely to exists when a big rare event occurs and that option trading should be clustered during the time of such events. However, if this is indeed the case, the likelihood to observe options with positive trading volume on two consecutive trading days is also higher at such times. The potential bias results from the fact that the JUMP and VOL factors are constructed to earn positive returns precisely when such events occur. As a consequence, the likelihood to obtain valid return observations for the JUMP and VOL factors is probably

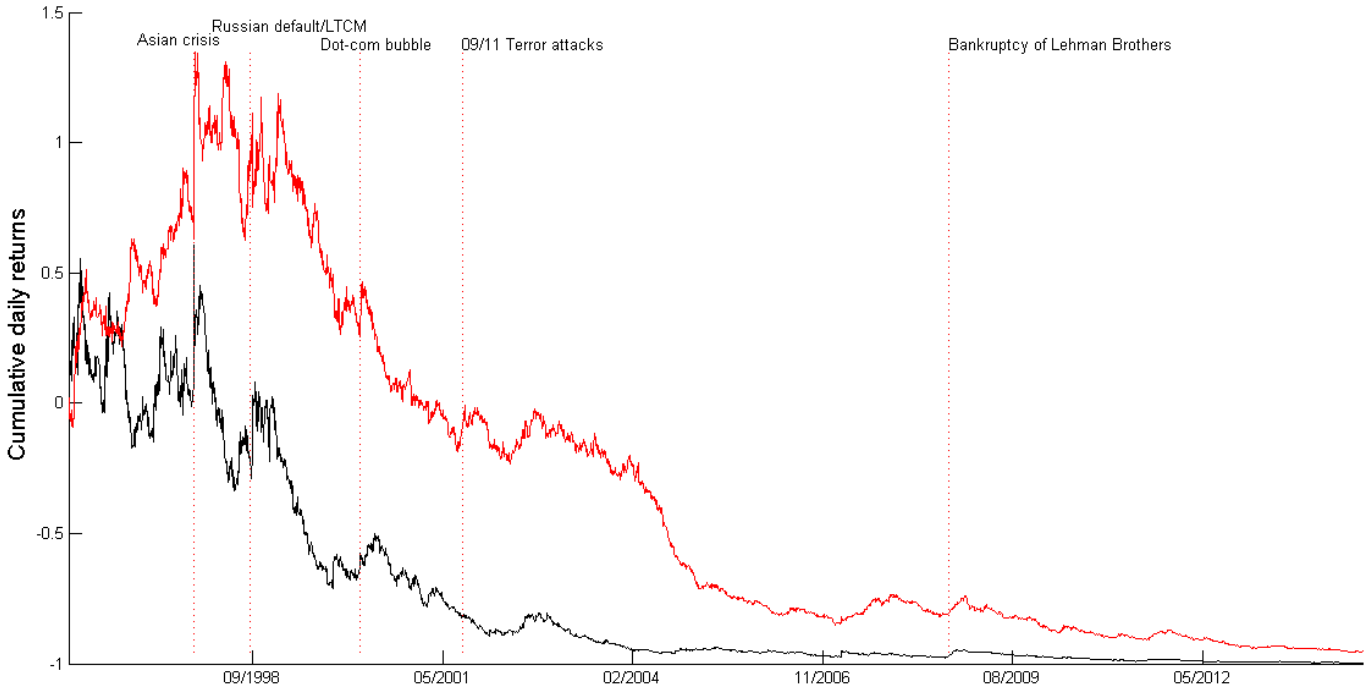

Figure 2: Cumulative daily returns on risk-factor-mimicking portfolios JUMP and VOL

Note: The black line shows the cumulative daily returns on the jump-risk mimicking portfolio JUMP (calculated from discrete returns) over the sample period from January 5th 1996 August 29th 2014. The red line shows the cumulative daily returns on the VOL factor over the same period. Vertical dashed lines indicate major events.

higher on days on which the two factors earn positive returns, which erroneously biases the average return upwards. Since positive returns on the constructed factor portfolios are frequently double-digit, this bias can potentially be substantial. This may partly explain the remarkably high positive average returns on the CAC40 and the SMI reported in Table 6. Note that for the two factors constructed from S&P 500 options, valid returns could be computed on more than 97.8% of all days in the sample so that the previously explained situation is unlikely to bias average returns on these two factors. For the VOL factors, the sign is not persistent across the five European indices and the mean return is never significantly different from zero.

Due to the frequent and often extended time gaps between valid factor return observations for the European indices, cumulative returns are significantly less meaningful than for the US factors. Nevertheless, for the sake of completeness, Figure A1 in the Online-Appendix shows the cumulative returns for the European indices.

As with the variance risk premium before, the returns on the two factor portfolios appear to stem from a non-normal distribution with substantial skewness and kurtosis that can partly be alleviated by using continuous returns instead of discrete. However, as noted by Coval and Shumway (2001) and indicated by the descriptive statistics for the JUMP factor shown above, logarithmic option returns are significantly lower than the raw option returns and this relocation is markedly more pronounced than for variance swap returns. Due to this, discrete returns will be used for the following analyses.

Empirical characteristics of risk-factor-mimicking portfolios: Determination of jump risk exposure

Because the validity of any conclusion drawn about the extent to which the variance risk premium reflects compensation for jump risk requires that the JUMP factor be able to measure jump risk accurately and not simply other priced factors, I examine whether positive returns on the JUMP factor coincide with jumps indicated by the non-parametric jump detection test of Lee and Mykland (2008) which allows to determine if a jump has occurred at a particular point of time. The underlying idea is that large price movements and associated returns can either be due to a discontinuous movement, i.e. represent a jump in the true sense, or simply be the result of a comparatively high spot volatility at the particular point of time. In order to reliably distinguish the two cases, the authors propose to standardize the return that is to be analyzed by the instantaneous volatility and derive the following test statistic, which determines whether a jump occurred from time $(t-1)$ to t.

$$\Lambda(t) \equiv \frac{log(S_t/S_{t-1})}{\widehat{\sigma(t)}}, \tag{14}$$

where $\widehat{\sigma(t)}$ represents the square root of the estimated bipower variation based on the previous W observations of

the spot price S_t at time t.

$$\widehat{\sigma(t)^2} \equiv \frac{1}{W-2} \sum_{j=t-W+2}^{t-1} (|\log(S_j/S_{j-1})| \, |\log(S_{j-1}/S_{j-2})|) \tag{15}$$

In constructing the test statistic and computing the realized bipower variation, I follow the recommendation of Lee and Mykland (2008) to use an optimal window size of $W = 16$ for daily observations. The test statistic follows approximately a normal distribution if there is no jump at the testing time, but becomes very large otherwise. A reasonable rejection region can be determined by comparing the test statistic with the usual region of maximums of the test statistic. If the test statistic is outside this usual region, it is probable that the examined price movement represents a jump in the true sense. Thus, the null hypothesis of no jump at a particular point of time can be rejected if $\frac{|\Lambda(t)|-C_n}{S_n}$ exceeds the critical value β^*, where $\beta^* = -\log(-\log(1-\alpha))$, $C_n = \frac{(2\log(n))^{1/2}}{c} - \frac{\log(\pi)+\log(\log(n))}{2c(2\log(n))^{1/2}}$, $S_n = \frac{1}{c(2\log(n))^{1/2}}$ and $c = \sqrt{2/\pi}$ for a given confidence level α and sample size n.

The analysis whether the JUMP factor truly captures exposure to price jumps is limited to the JUMP factor based on S&P 500 options. The reason is that the JUMP time series for the S&P 500 index is by far the longest and can therefore be expected to offer the most meaningful results.

Figure 3 shows the time series of daily returns on the JUMP factor. The vertical dashed lines indicate the occurrence of jumps determined by the applied jump detection test. Based on a confidence level of $\alpha = 5\%$, the jump test identifies 66 days in the time series of the S&P 500 index on which jumps occur. As can be clearly seen in the figure, jumps in the index are usually accompanied by positive returns on the JUMP factor. In particular, the returns on the JUMP factor are positive for 62 of the 66 days on which a jump is detected with an average daily return of +8.39% on jump days compared to an average daily return of -0.19% on days without a jump. Even though this is a first indication that the JUMP factor indeed captures the effect of price jumps, it is also apparent in Figure 3 that returns on the JUMP factor are positive on many days on which the test does not detect a jump. In this context note that the asymptotic arguments used for setting up the test statistic and the rejection region strongly depend on the assumption that the time period between consecutive observations converges to 0. As a consequence, the use of daily data naturally inhibits the detection of jumps. Lee and Mykland (2008) point out that only 2% of jumps can be detected using daily return data. Thus, the test results should only be regarded as an additional indication to support the theoretical argument for the JUMP factor's exposure to price jumps given above, and not as an independent attempt to prove this exposure.

To further link the returns on the JUMP factor to price jumps, daily JUMP factor returns that occur on days that are no jump days but directly follow such days are compared with the corresponding returns on days that neither immediately follow nor are jump days. The average daily return on days that immediately follow a jump day is -0.87% compared to -0.18% on other days. A regression of the daily JUMP factor returns on non-jump-days on an indicator variable that is one when the return occurs on a day that immediately follows a jump day but is not a jump day itself, and zero when the return occurs on a day that does not follow a jump day and is not a jump day itself, obtains a t-statistic of -1.8595 for the indicator variable. Thus, it seems that investors require a higher compensation for bearing the risk to which the JUMP factor offers exposure, presumably jump risk, immediately after a price jump has occurred. Conceptually, this observation is plausible with regard to an investor who reevaluates the jump probability and deems the occurrence of a further jump more likely after a jump has occurred. Such a behavior would also be consistent with the empirical evidence that financial markets exhibit strong self-excitation over time (e.g. Aït-Sahalia et al. (2015a)), i.e. that jumps are often followed by further jumps or substantial price movements. Possible reasons that could explain this pattern include credit and liquidity shocks and margin calls that are caused by preceding jump events and then trigger further jumps (Aït-Sahalia et al. (2015a)). An alternative explanation could be investors' increased desire to protect against further market drops as in the model by Bates (2008) where wealth redistributions due to preceding market drops increase average crash aversion.

All in all, the return behavior appears to be reconcilable with the notion that the JUMP factor indeed captures exposure to jump risk, which is therefore assumed for the remainder.

Compensation for jump risk and diffusive risk in variance swap returns

For the following analyses, I follow Cremers et al. (2015) and treat the returns on the two risk factors directly as excess returns that can be used in the regressions. In order to compare variance swap returns and returns on the JUMP and VOL factors over the same time horizon, the cumulative factor returns over 22 trading days are used in the following regressions. Note that the requirement to have returns on 22 consecutive trading days substantially reduces the numbers of available observations compared to those reported in Table 6. For example for the SMI there is no single period with valid returns over 22 consecutive trading days.

Table 7 shows the results of overlapping regressions of the LVRP on the JUMP and VOL factors.

The obtained R^2s are generally high and exceed 60% for five of the seven indices with more than 90% for the CAC40. As expected, the coefficient on the VOL factor is positive and significant at the 1% level for most regressions, which confirms that exposure to changes in systematic variance is heavily priced not only in variance swaps on indices but also in such on individual stocks. Moreover, consistent with prior studies which find that a substantial portion of the observed variance risk premium represents compensation for jump risk (e.g. Todorov (2010), Bollerslev and Todorov (2011)), the JUMP factor is significant at the 1% level for all underlyings

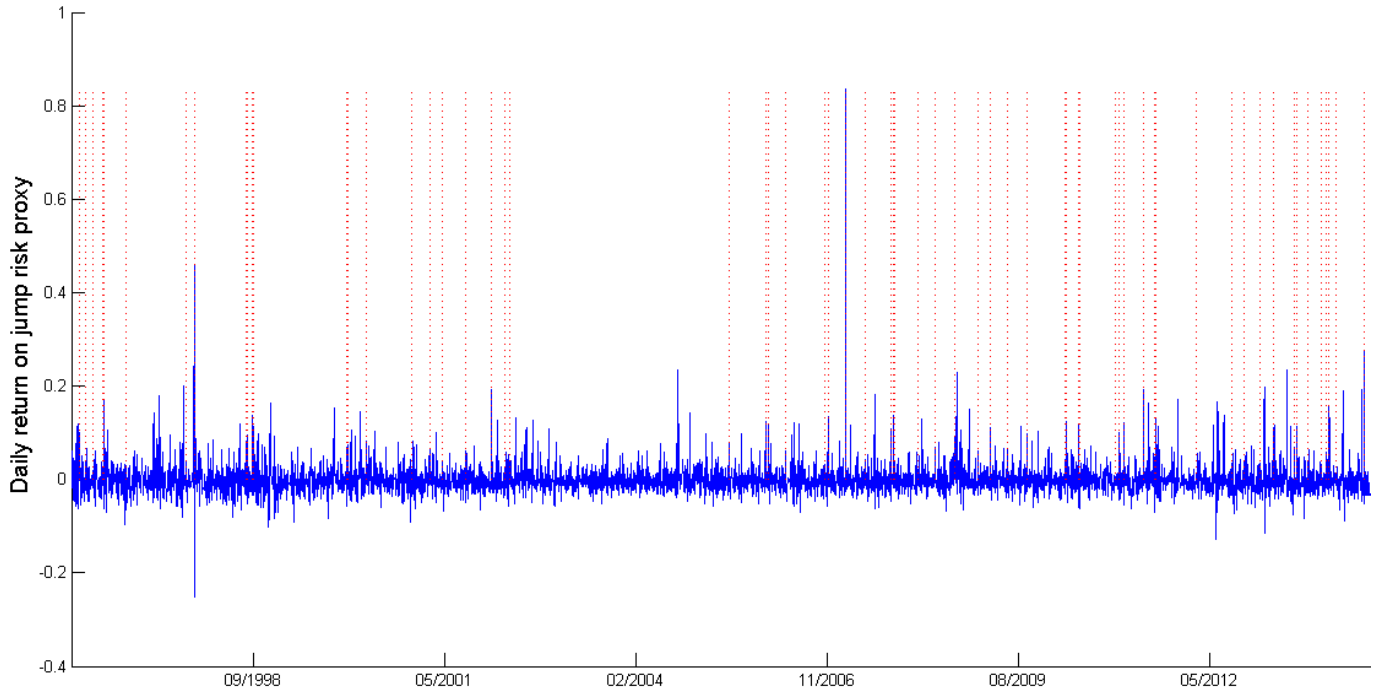

Figure 3: Daily returns on jump risk proxy

Note: The figure shows the time series of daily discrete returns on the jump risk factor JUMP over the period from January 5th 1996 August 29th 2014. Vertical dashed lines indicate the occurrence of jumps based on the jump detection test of Lee and Mykland (2008) using a confidence level of $\alpha = 5\%$.

except for Tesla Motors and Facebook. The obtained sign of the coefficient is also consistent with the notion that variance swaps offer protection against jumps. Jumps increase the realized variance, and if their impact is sufficiently large, payoffs on variance swaps are positive just as are payoffs on the JUMP factor. The fact that both factors are jointly significant suggests that investors clearly distinguish between pure volatility risk and jump risk and that each of these risks commands a separate premium that contributes to variance swap returns. The average contributions (in absolute terms) of the compensation for jump risk and diffusive risk to variance swap returns appear to be of comparable magnitude. Multiplying the respective regression coefficient by the average risk factor return and taking the average across all underlyings results in a mean contribution of -2.95% for the JUMP factor and -2.34% for the VOL factor. However, even though the JUMP and VOL factors should directly proxy for the two risk factors most commonly thought to be priced in options, the alpha is still highly significant and persistently negative for all underlyings except the CAC40, for which variance swap returns are fully explained by the two factors.

In order to further investigate the explanatory power of the JUMP and VOL factors in the presence of traditional factors, I augment the previous regressions with the Fama and French (1993) factors. The results are shown in Table 8. First, even in the presence of the three traditional factors, the coefficients on both the JUMP and VOL factors remain positive and significant at the 1% level for the majority of regressions, which suggests that the two factors contain information that are not already captured by the Fama and French (1993) factors.

Second, for the majority of regressions, the absolute magnitude of the SMB coefficient as well as the total number of regressions for which it is significant is smaller than in Table 5 after the inclusion of the two risk-factor-mimicking portfolios, which points to a link between the SMB factor and either jump or volatility risk or both. Individual regressions with the Fama and French (1993) factors and the VOL and JUMP factor separately show that only the JUMP factor leads to this reduction in the number of significant SMB coefficients. In contrast, for regressions with the VOL factor and without the JUMP factor, the number of significant coefficients on the SMB factor does not change. In this context, the presence of priced jump risk in variance swap returns may also help to rationalize the negative and often significant coefficient on the SMB factor in previous regressions with the Fama and French (1993) factors (c.f. Table 5), that is also documented by Carr and Wu (2009). A possible relation between the SMB factor and jump risk can be substantiated by recent findings of Bollerslev et al. (2015) who examine how individual equity prices are affected by continuous and discontinuous market price movements. For this purpose, they calculate separate betas to measure the continuous and discontinuous comovements. Portfolio sorts on the discontinuous betas lead to statistically significant positive return spreads between the High-Low-portfolio but less so for the continuous beta. Moreover, their results show that particularly smaller firms and stocks with lower book-to-market ratio tend to have higher

Table 7: Regression of variance swap returns on returns on jump-risk mimicking portfolio JUMP and volatility-risk mimicking portfolio VOL

Note: Entries show the OLS estimates and t-statistics (in parentheses) from regressions of continuously compounded 22 day variance swap returns on the 22-day cumulative returns on the jump-risk mimicking portfolio JUMP and the VOL factor. t-statistics are based on standard errors that are adjusted for serial correlation according to the method of Newey and West (1987) with a lag length of 22 days. N denotes the number of total observations.

Underlying	alpha		JUMP		VOL		R^2	N
S&P 500	-0.433	(-16.574)	2.589	(15.635)	2.976	(8.739)	62.93%	2881
Dow Jones Industrial	-0.396	(-11.696)	2.322	(11.512)	2.999	(7.508)	57.72%	1904
NASDAQ100	-0.282	(-9.686)	2.008	(9.625)	2.080	(5.914)	45.08%	2637
Alcoa	-0.233	(-4.142)	1.106	(3.718)	1.376	(1.541)	19.01%	703
Altria (Philip Morris)	-0.338	(-6.209)	0.740	(2.275)	2.480	(3.486)	8.42%	1658
Amazon	-0.222	(-5.173)	1.099	(3.978)	1.608	(2.966)	12.33%	2023
American Express	-0.137	(-3.592)	1.905	(9.443)	2.714	(5.134)	39.49%	1616
Amgen	-0.289	(-7.176)	1.421	(6.846)	1.618	(2.796)	20.04%	1933
Analog Devices	-0.162	(-3.072)	1.325	(5.780)	1.203	(1.802)	18.08%	536
Apple	-0.299	(-7.656)	1.080	(3.976)	1.306	(2.994)	15.39%	1962
Bank of America	-0.152	(-2.652)	1.770	(5.213)	1.210	(2.057)	23.60%	1236
Boeing	-0.213	(-6.337)	1.090	(5.506)	1.390	(2.807)	17.23%	1778
Cisco	-0.197	(-4.056)	1.250	(4.704)	2.699	(4.993)	21.55%	1436
Exxon Mobil	-0.253	(-6.377)	1.557	(6.869)	2.946	(5.880)	37.04%	1565
Facebook	-0.340	(-4.501)	0.345	(0.696)	-3.853	(-2.098)	8.67%	486
General Electric	-0.169	(-4.821)	1.840	(11.456)	1.932	(3.739)	35.11%	1446
Home Depot	-0.265	(-7.672)	1.400	(7.201)	2.308	(5.750)	27.27%	1537
IBM	-0.295	(-8.682)	1.447	(7.250)	2.430	(6.567)	24.33%	2298
Johnson & Johnson	-0.330	(-7.060)	2.101	(6.843)	2.996	(4.689)	34.08%	1058
McDonald's	-0.384	(-9.678)	0.839	(3.929)	2.514	(5.009)	22.97%	1286
Merck	-0.237	(-5.920)	2.012	(8.218)	1.679	(3.194)	23.77%	1663
Metlife	-0.126	(-2.836)	1.669	(4.114)	3.286	(6.193)	47.67%	900
Microsoft	-0.208	(-5.128)	1.213	(5.149)	1.200	(2.482)	14.99%	1863
Monsanto	-0.244	(-6.749)	1.319	(6.487)	2.094	(4.115)	30.19%	1217
Nike	-0.249	(-5.134)	1.165	(4.071)	2.882	(5.091)	21.57%	1241
Pfizer	-0.223	(-5.263)	2.009	(8.171)	2.010	(4.329)	26.51%	1139
Procter & Gamble	-0.357	(-6.448)	1.806	(5.367)	2.757	(6.146)	29.64%	1393
Starbucks	-0.229	(-4.446)	0.729	(3.455)	1.565	(2.006)	9.45%	1316
Tesla	-0.277	(-3.817)	0.096	(0.214)	0.758	(0.781)	0.74%	701
Valero	-0.151	(-3.473)	1.086	(3.482)	0.905	(1.534)	14.11%	1607
Verizon	-0.257	(-6.726)	1.075	(4.184)	2.136	(3.336)	22.76%	1326
WalMart	-0.317	(-7.693)	1.006	(5.173)	2.372	(5.218)	19.25%	1584
AEX	-0.382	(-8.334)	2.496	(4.356)	4.337	(5.929)	61.52%	523
CAC40	0.013	(1.141)	6.140	(30.556)	0.438	(2.296)	90.14%	22
DAX	-0.267	(-6.701)	3.337	(9.152)	3.544	(6.155)	66.22%	940
Euro Stoxx 50	-0.295	(-6.519)	3.266	(10.115)	2.105	(4.067)	64.26%	485
SMI	-	-	-	-	-	-	-	-

discontinuous betas.

This implies that small stocks have a more pronounced tendency to co-jump with the market than larger stocks. Ang and Chen (2002) also document that correlation asymmetries are stronger for smaller stocks and conclude that such stocks are more exposed to common downward movements with the market. Thus, the SMB factor should have a tendency to perform well in good times but less so in bad times when negative jumps occur and variances usually increase. This relation is also supported by the negative correlation of -0.2468 (p-Value 0.0003) between the returns on the JUMP factor and on the SMB factor for the US. As a consequence variance swap returns appear to be negatively correlated with returns on the SMB factor due to their differential exposure to jumps.

Third, even though the coefficient on the excess market return is still negative and significant at the 1% level for almost all regressions, its magnitude is substantially smaller by an amount of as much as approximately one half compared to the previous regressions without the two additional factors. This finding suggests that information contained in the market excess return is also captured by the JUMP and VOL factors. The negative correlations of -0.2643 (p-Value 0.0003) between returns on the US market portfolio and on the US JUMP factor and of -0.3091 (p-Value 0.0000) between the US market excess return and returns on the VOL fac-

Table 8: Regression of variance swap returns on Fama-French and JUMP and VOL factors

Note: Entries report the OLS-estimates and t-statistics (in parentheses) of non-overlapping regressions of 22-day continuously compounded variance swap returns on the corresponding excess return on the market portfolio and the two additional factors, SMB and HML, identified by Fama and French (1993), as well as on the jump-risk mimicking portfolio JUMP and the VOL factor. For the US, the market excess returns is proxied by the 22-day excess return on the value-weighted portfolio of all stocks traded at the NYSE, NASDAQ, and AMEX. For the European indices, the market excess return refers to the excess return on a proxy for a European market portfolio taken from Kenneth French's website. t-statistics are not adjusted for serial correlation. N denotes the total number of observations.

Underlying	alpha		market		SMB		HML		JUMP		VOL		R^2	N
S&P 500	-0.434	(-14.412)	-4.534	(-6.140)	-0.573	(-0.494)	-0.628	(-0.629)	2.083	(10.699)	2.307	(5.834)	69.19%	174
Dow Jones Industrial	-0.383	(-11.080)	-6.253	(-6.210)	2.503	(1.749)	2.196	(1.723)	1.695	(7.144)	2.577	(5.388)	68.20%	125
NASDAQ100	-0.287	(-8.419)	-4.151	(-5.009)	-0.936	(-0.974)	-0.028	(-0.030)	1.662	(8.459)	1.527	(3.300)	55.86%	172
Alcoa	-0.249	(-3.992)	-6.445	(-3.286)	-1.916	(-0.972)	3.493	(2.066)	0.599	(1.885)	-0.015	(-0.013)	47.96%	45
Altria (Philip Morris)	-0.349	(-4.789)	-1.162	(-0.929)	-1.723	(-0.861)	-3.483	(-2.336)	0.474	(1.367)	2.307	(3.224)	13.77%	124
Amazon	-0.182	(-3.336)	-0.647	(-0.428)	-2.736	(-1.110)	-4.531	(-2.021)	0.929	(2.135)	1.229	(1.563)	15.10%	138
American Express	-0.103	(-2.143)	-2.510	(-2.751)	0.563	(0.421)	0.585	(0.511)	1.544	(5.193)	2.188	(3.263)	41.00%	110
Amgen	-0.281	(-5.438)	-2.268	(-2.522)	-0.787	(-0.683)	-1.252	(-0.763)	1.279	(4.795)	0.978	(1.229)	27.44%	129
Analog Devices	-0.091	(-1.192)	-1.722	(-1.543)	-4.996	(-2.669)	-1.396	(-0.721)	1.122	(2.538)	-0.415	(-0.406)	38.24%	33
Apple	-0.287	(-6.129)	-3.790	(-4.425)	-1.034	(-0.821)	-0.045	(-0.035)	0.992	(2.809)	0.59	(1.056)	27.82%	126
Bank of America	-0.074	(-1.012)	-5.260	(-4.799)	-1.612	(-1.055)	-3.760	(-2.494)	1.294	(3.131)	-0.325	(-0.435)	43.48%	76
Boeing	-0.167	(-4.037)	-2.840	(-2.611)	-1.453	(-0.770)	-0.483	(-0.341)	0.763	(3.468)	0.464	(0.634)	23.27%	119
Cisco	-0.204	(-4.020)	-4.237	(-4.272)	1.164	(0.643)	0.577	(0.489)	1.022	(2.795)	1.676	(2.077)	30.98%	97
Exxon Mobil	-0.260	(-6.512)	-4.302	(-3.502)	1.113	(0.560)	-0.287	(-0.189)	1.136	(4.657)	2.733	(3.625)	49.83%	105
Facebook	-0.367	(-6.826)	-4.448	(-1.445)	5.736	(1.268)	6.62	(2.260)	0.967	(2.764)	-5.636	(-3.168)	31.27%	23
General Electric	-0.174	(-3.874)	-2.413	(-3.576)	-1.462	(-0.976)	-0.441	(-0.393)	1.579	(6.287)	1.212	(1.987)	38.33%	96
Home Depot	-0.233	(-4.689)	-1.955	(-3.001)	-1.654	(-1.420)	0.705	(0.727)	1.229	(4.033)	2.375	(4.416)	39.90%	99
IBM	-0.291	(-7.275)	-2.786	(-3.708)	-1.852	(-1.998)	-0.668	(-0.527)	1.199	(5.039)	1.874	(3.836)	36.87%	155
Johnson & Johnson	-0.364	(-6.360)	-1.833	(-1.666)	-3.628	(-2.646)	-0.281	(-0.144)	1.402	(2.328)	2.400	(2.483)	36.04%	70
McDonald's	-0.388	(-8.289)	-2.854	(-1.795)	1.317	(0.540)	1.720	(0.648)	0.523	(2.112)	2.485	(3.278)	35.24%	80
Merck	-0.240	(-4.310)	-1.257	(-1.552)	-0.868	(-0.691)	1.810	(1.576)	1.554	(3.585)	1.757	(2.074)	25.50%	110
Metlife	-0.102	(-2.157)	-4.465	(-2.579)	3.116	(1.067)	0.943	(0.668)	1.491	(3.483)	2.784	(3.400)	66.16%	60
Microsoft	-0.180	(-3.637)	-3.385	(-3.505)	-2.517	(-1.872)	-0.641	(-0.540)	0.697	(2.129)	0.317	(0.523)	24.78%	126
Monsanto	-0.214	(-5.027)	-1.466	(-1.378)	-1.172	(-0.500)	-3.011	(-1.791)	0.938	(3.132)	2.172	(3.736)	38.92%	84
Nike	-0.193	(-3.050)	-2.252	(-1.344)	-1.759	(-0.666)	-2.541	(-0.859)	0.659	(1.435)	2.956	(5.306)	27.76%	74
Pfizer	-0.186	(-2.791)	-1.688	(-1.597)	-1.047	(-0.740)	-2.411	(-2.147)	1.858	(5.546)	1.658	(2.623)	34.09%	77
Procter & Gamble	-0.356	(-5.129)	-2.466	(-2.206)	-6.317	(-2.306)	2.116	(1.360)	0.883	(3.116)	2.084	(4.415)	42.56%	87
Starbucks	-0.232	(-4.765)	-3.776	(-3.289)	0.428	(0.223)	-1.412	(-0.671)	0.554	(2.293)	0.159	(0.118)	17.12%	88
Tesla	-0.337	(-4.944)	-0.662	(-0.292)	6.570	(1.531)	-1.780	(-0.274)	-0.088	(-0.183)	1.596	(1.368)	7.45%	43
Valero	-0.151	(-2.767)	-4.019	(-2.612)	2.854	(1.079)	0.618	(0.302)	0.689	(2.030)	0.007	(0.008)	15.86%	112
Verizon	-0.225	(-4.627)	-2.223	(-1.780)	-0.296	(-0.142)	-0.163	(-0.094)	0.829	(2.907)	2.582	(2.981)	38.86%	79
WalMart	-0.320	(-5.841)	-3.246	(-4.019)	-3.401	(-2.482)	0.123	(0.113)	0.808	(2.757)	0.665	(0.859)	29.00%	106
AEX	-0.311	(-5.383)	1.533	(0.668)	-0.007	(-0.281)	-0.054	(-1.943)	3.028	(4.220)	5.531	(3.015)	82.61%	21
CAC40	-	-	-	-	-	-	-	-	-	-	-	-	-	0
DAX	-0.257	(-5.038)	-3.656	(-2.718)	-0.008	(-0.303)	-0.014	(-0.600)	2.315	(4.904)	2.127	(2.437)	73.11%	40
Euro Stoxx 50	-0.212	(-2.876)	-4.821	(-1.770)	0.012	(0.219)	-0.039	(-1.636)	1.842	(1.520)	2.399	(0.951)	73.55%	21
SMI	-	-	-	-	-	-	-	-	-	-	-	-	-	0

tor for the US further support this view. For one thing, the negative correlations are not surprising since the JUMP factor usually earns positive returns at times when substantial (mostly negative) market moves occur and such moves are often associated with elevated levels of volatility, which in turn affect returns on the VOL factor. However, the underlying economic rationale to which this association can probably be attributed is the notion that a substantial portion of the equity risk premium itself represents compensation for jump risk and volatility risk (e.g. Bollerslev and Todorov (2011)). The intuitive explanation for this relation is that an increased volatility adversely affects investors' investment opportunity sets and that jumps may be associated with tail risk.[14] If investors then feel that the likelihood of a tail event or return volatility increases, they may require a higher return, which leads to an immediate, often substantial, reduction in asset prices. This mechanism is quite similar to the volatility feedback effect explained before. A possible explanation for why elevated volatility levels or perceived fear of extreme events can have important impacts on asset prices and thus on market excess returns is that such factors are suspected to affect aggregate economic output. This is the so-called uncertainty shock channel. For instance, Bloom (2009) finds evidence that uncertainty shocks have substantial effects on the aggregate economy, leading to significant drops in production, output, and productivity growth since higher uncertainty causes firms to delay investments and hiring. While Bloom (2009) relates uncertainty to volatility, Gourio (2012) and Kelly and Jiang (2014) find that perceived tail risk can have similar effects on aggregate economic activity. The changing magnitude of the coefficient on the market excess return after the introduction of the VOL and JUMP factor can therefore be regarded as a further indication that the equity risk premium is linked to jump or tail risk as well as volatility risk.

Finally, even though the alpha coefficient is slightly less negative when the JUMP and VOL factors are included as additional explanatory variables, their overall influence on its magnitude is negligible and it remains significant in all regressions. While it presumably is comprehensible that commonly used risk factors such as those proposed by Fama and French (1993) cannot satisfactorily explain the returns on variance swaps when overall market variance and jump risk constitute independently priced risk factors, it is astonishing that adding the two constructed factor portfolios does not appear to perceptibly reduce the alpha. Especially with regard to the fact that the coefficients on the JUMP and VOL factors behave as one would expect them to do, so that it can reasonably be assumed that these two factors are indeed able to effectively capture the isolated compensation for jump and volatility risk, this is intriguing. Moreover, even if one considers the possibility that a certain portion of single stock variance swap returns represents compensation for the risk of common changes in idiosyncratic variance, as for example suggested by Gourier (2015), this is unlikely to be the case for highly diversified stock indices such as the S&P 500 for which the alpha is also significant. Consequently, it appears that the pricing of variance swaps is either inefficient, allowing to earn significant abnormal returns that appear to be unrelated to systematic risk factors or that sources of systematic risk other than jump and volatility risk are priced in variance swap rates.

Robustness considerations

Since the assertion that financial assets are inefficiently priced should be made with caution only, it is imperative to take into account possible misspecifications and consequences of transformations of variance swap returns that may have an impact on the results. As previously outlined, the logarithmic version of the variance risk premium, defined as $LVRP_{t,T} = \log(RV_{t,T}/E_t^{\mathbb{Q}}[RV_t, T])$, is introduced in order to account for the substantial kurtosis and positive skewness of raw variance swap returns – particularly driven by excessively high, positive returns on long variance swap positions during the recent financial crisis – and thereby make the distribution more Gaussian. The underlying rationale for this transformation is to make the regression residuals more normal. Even though this transformation is also applied by other studies (e.g. Carr and Wu (2009)) and the use of log returns in the context of stock return applications is widely accepted since deviations between continuously compounded returns and discrete returns are typically negligible for small price differentials, its application to variance swap returns comes with potential drawbacks. While both, skewness and kurtosis can indeed be mitigated through the log transformation, the often substantial divergence between realized variance and the variance swap rate can lead to considerable deviations between raw returns ($RV_{0,T}/E_t^{\mathbb{Q}}[RV_{0,T}] - 1$) and their logarithmic version. As a consequence, the logarithmic transformation tilts the mean of the variance swap return distribution disproportionately into the negative domain, which may contribute to the persistently negative alpha.

In order to investigate the impact of the logarithmic transformation on reported results, several robustness regressions are performed. Since the aim is only to provide qualitative insights, the corresponding results can be found in the Online-Appendix.

Table A3 in the Online-Appendix shows the results from robustness regressions of raw 22-day variance swap returns on monthly realizations of the three Fama and French (1993) factors and the two risk factor portfolios JUMP and VOL over the entire sample horizon from January 1996 to July 2015. For raw variance swap returns, the alpha is consistently less negative and significant at the 5%-level for only 12 underlyings including the indices, compared to 33 significant alphas for the logarithmic specification.

In order to further emphasize the influence of the logarithmic transformation, the sample is divided into two subsamples. The first subsample is restricted to cover only the time period from January 1996 to December 2007 and there-

[14] For instance, Bollerslev and Todorov (2011) find that the magnitude of the left jump tail which captures fears of dramatic market declines significantly exceeds that of the right-jump tail which is associated with substantial market appreciations.

fore to exclude the time period of unprecedented uncertainty during and following the financial crisis. Since extremely positive as well as extremely negative variance swap returns predominantly occur in later periods, the effect of the logarithmic transformation can be expected to be comparatively moderate in this subsample. Note that this subsample only comprises the US underlyings since the time series for the European underlyings either start later or it is not possible to obtain at least ten valid observations for all regressors (this is the case for the DAX). The second subsample is restricted to cover the time period from January 2008 until July 2015 during which the qualitative and quantitative deviations between raw returns and their logarithmic version can be expected to be more pronounced. Results for the first and second subsample for the same regression specifications as before are reported in Tables A4 and A5 in the Online-Appendix.

Comparing the results for the entire sample and the first subsample, the results differ and deviations between the two specifications are less pronounced for the latter. For the first subsample, the number of negative and significant alphas for the logarithmic specification reduces substantially from 33 to only 15 underlyings whereas the specification with raw returns produces significantly negative alphas for eight underlyings. For both specifications, however, the alphas remain significant for the indices. As expected, qualitative and quantitative differences are more pronounced for the second subsample since the most extreme returns – positive and negative – occur in this period. Regressions with continuously compounded returns again produce 32 significantly negative alphas compared to 12 for regressions with raw returns.

These robustness regressions clearly highlight the importance of the chosen specification of the variance risk premium as well as of the chosen time period and provide evidence that the two specifications might lead to qualitatively different conclusions for certain underlyings when the objective is to determine whether variance swap returns are completely explained by a given set of risk factors. However, even though the number of significant alphas is obviously lower when raw variance swap returns are used, none of the applied regression models is able to fully explain these returns for all underlyings over the entire time horizon or subsets of it. Thus, even though variance swap returns clearly include compensation for jump and volatility risk, these risks are not the only factors that appear to influence the pricing of options and variance swaps, which points toward alternative or additional explanations such as for example model uncertainty (e.g. Drechsler (2013)) or demand-based explanations (e.g. Gârleanu et al. (2009)).

5.2.4. Model uncertainty and uncertainty about macroeconomic fundamentals

In this final section, variance swap returns are related to model uncertainty and uncertainty about macroeconomic fundamentals. The purpose is explicitly not to explain the variance risk premium through risk factors but rather to shed some light on one of the factors that could contribute to the time-varying differences in investors' perception of variance risk or jump risk that are presumably associated with the substantial fluctuations of the variance risk premium over time (c.f. Figure 1) and to assess whether a measure of model uncertainty could be helpful in explaining variance swap returns in a richer setting than that used here.

The variable to be considered in this setting is motivated by the work of Drechsler (2013) who considers how the variance risk premium is affected when investors face model uncertainty. Due to the model uncertainty, the investor is not only uncertain about a possible outcome but also about the data generating process and thus evaluates his decisions under the "worst case" scenario, in which he particularly overestimates the magnitude and frequency of jumps. Since the investor knows that realized variance will be high if the worst case scenario manifests, his willingness to pay for variance swaps will also be high when model uncertainty is high. Therefore, a direct implication of the model is that the size of the variance risk premium will fluctuate and reflect investors' degree of model uncertainty. In particular, this relation is under consideration in the following.

In order to measure model uncertainty, I follow Drechsler (2013) and use the cross-sectional dispersion (DispGDP), i.e. the variance of implied quarterly growth rates of the nominal gross domestic product (GDP). Consequently, higher values are associated with greater model uncertainty.

The necessary data is taken from the Survey of Professional Forecasters to which the Federal Reserve Bank of Philadelphia provides access.[15] The Survey of Professional Forecasters contains forecasts by individual forecasters for the quarterly level of nominal GDP in billions $, based on a seasonally adjusted annual rate.

The forecasted growth rate is based on the forecasted level of nominal GDP for the current quarter and the realized level of nominal GDP in the previous quarter. In order to mitigate the effect of outliers, dispersion figures which are more than two standard deviations from the sample mean are removed. Drechsler (2013) argues that information asymmetries can be expected to play only a minor role in the context of forecasting macroeconomic quantities such as GDP because essential information are publicly available. As a consequence, the dispersion of GDP forecasts can naturally be interpreted as the result of different economic models.

Since the individual GDP forecasts are made each quarter based on an adjusted annual growth rate whereas the variance swap returns refer to a period of only 22 trading days, I assume that the DispGDP measure does not only reflect model uncertainty for a particular day but reflects uncertainty around the time the forecasts are made, i.e. at the beginning of a quarter. As a consequence the DispGDP measure at the beginning of a given quarter is matched with the variance swap return over the 22 trading day period starting on the day on which the DispGDP measure is computed. If no such return could be computed on this specific day, I

[15] The internet address is: https://www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters/

use the next available return calculated on a later day but not later than three calendar days after the day the DispGDP measure is computed. The log of the realized variance over the 22 trading day period ending immediately before the period starts over which the variance swap returns are computed is included as an additional control variable because other studies (e.g. Bollerslev et al. (2011)) document a significantly positive relationship between previously realized volatility and risk-aversion as measured by the variance risk premium. The reason for taking the log of realized variance is that variances have right-skewed distributions whereas logarithmic variances tend to have near Gaussian distributions, which may improve predictions from linear models (Bekaert and Hoerova (2014)).

Since both variables, DispGDP and the preceding variance are measures of uncertainty, including them simultaneously in the regressions could lead to flawed inferences if both measures are highly correlated and convey widely identical information. However, for 23 of the 32 underlyings, the absolute value of the correlation between DispGDP and the previously realized variance is below 0.15 with a maximum correlation of 0.47 and an average value across all underlyings of 0.0306. Thus, the two measures appear to convey information about relatively unrelated sorts of uncertainty.

Table 9 shows the results from regressions of the log variance risk premium on the log of the realized variance during the preceding 22 trading days and DispGDP.

The coefficient on the DispGDP measure is negative for all but two underlyings and significant for some of them albeit not for all suggesting that model uncertainty is indeed associated with a more negative realized variance risk premium. The sign of the coefficient is therefore consistent with the notion that investors are willing to pay more for an instrument that offers insurance against undesirable states when the degree of model uncertainty, i.e. uncertainty about the distribution of possible outcomes, is higher. Even though the regression specification applied here is of course not directly comparable with the model specification of Drechsler (2013), the results are in line with his argumentation that model uncertainty and the associated uncertainty or disagreement about macroeconomic fundamentals may help to explain the magnitude and often substantial fluctuations in the observed variance risk premium. When turning to the realized variance over the preceding 22 trading days, the mostly positive sign on the coefficient contradicts the expected relation that higher variance could be associated with higher risk-aversion in the aftermath. However, the coefficients are mostly insignificant.

The chosen regression specification in this section is relatively parsimonious. As a consequence the results have limited meaning only. However, especially in consideration of the fact that the two factor portfolios that capture jump and volatility risk are not able to fully explain variance swap returns for several underlyings suggests that these returns may reflect compensation for further systematic risks. With regard to this, the findings in this section suggest that model uncertainty may be a potential candidate for such a risk factor whose influence on the pricing of options, and particularly of variance swaps, could be assessed with more advanced econometric techniques in the future.

6. Conclusion

In this work, I apply a relatively model-free method to extract the expected variance under the risk-neutral measure from prices of traded options and to quantify the realized variance risk premium.

Consistent with prior studies, I find a persistently negative variance risk premium for all stock indices. Moreover, the average return on a long single stock variance swap investment is also negative for all considered underlyings even though the absolute magnitude is often smaller than for the indices and returns exhibit a substantially greater variation. The finding of a negative variance risk premium for single stocks is in line with the results of some studies but contradicts those of others. However, these differential findings may, at least in part, be due to the application of different sample selection criteria since any additional factors and circumstances that systematically affect the cross-section of option prices, such as for example illiquidity discounts, will necessarily be reflected in the estimates of the risk-neutral expected variance and the variance risk premium.

All in all, the persistently negative premium suggests that investors regard variance swaps or the associated option positions as valid instruments to insure against undesirable states and are therefore willing to accept negative average returns on them. Even though the results suggest that continuously selling variance swaps would probably have been profitable for most underlyings over the entire sample period, an analysis of variance swap returns during the last quarter of 2008 reveals that the short side is exposed to considerable risks and the losses can be of excessive magnitude. The maximum returns on synthetic variance swaps during this period amount to several hundred percent for many underlyings.

Apart from quantifying the variance risk premium, the objective of this work is to assess whether return variance constitutes an independently priced risk factor for which the variance risk premium offers compensation.

The analysis reveals that the variance risk premium on the US indices is clearly linked to the return variance on the market portfolio and is more negative when the return variance of the index exhibits stronger covariation with that of the market portfolio. This is indicative of priced market variance risk. In contrast, a similar relation is not discernible for single stocks. However, excess returns on single stock variance swaps can often be fully explained by a systematic variance risk factor proxied by returns on variance swaps with the S&P 500 as underlying which suggests that returns on index as well as on single stock variance swaps are, to a considerable extent, driven by the same factors.

Further analyses show that the classical capital asset pricing model as well as the Fama and French (1993) three-factor-model are not able to explain variance swap returns and leave a significantly negative alpha, which further supports the notion of an independently priced market variance risk factor.

Table 9: Regressions of variance swap returns on a measure of model uncertainty

Note: Entries report the OLS-estimates and t-statistics (in parentheses) of non-overlapping regressions of 22-day continuously compounded variance swap returns on a DispGDP and the realized variance over the preceding 22 trading days (pastVariance). t-statistics are not adjusted for serial correlation. N denotes the total number of observations.

Underlying	const		DispGDP		pastVariance		R^2	N
S&P500 Index	-0.391	(-4.599)	-0.385	(-3.927)	2.765	(1.316)	15.33%	70
Dow Jones Industrial	-0.358	(-3.547)	-0.527	(-2.731)	2.876	(0.815)	15.77%	42
NASDAQ100	-0.210	(-2.680)	-0.151	(-1.935)	1.057	(2.189)	8.19%	68
Alcoa	-0.742	(-3.272)	0.409	(1.301)	2.234	(1.538)	22.46%	19
Altria (Philip Morris)	-0.531	(-3.450)	-0.083	(-0.576)	1.632	(1.279)	3.77%	43
Amazon	0.166	(1.395)	-0.064	(-0.629)	-0.024	(-0.445)	0.46%	51
American Express	-0.083	(-0.915)	-0.305	(-2.591)	0.494	(3.560)	13.13%	46
Amgen	-0.099	(-0.804)	-0.484	(-2.726)	0.270	(0.605)	7.25%	46
Analog Devices	0.177	(0.724)	-0.490	(-1.556)	-0.011	(-0.189)	8.89%	15
Apple	-0.231	(-3.075)	0.001	(0.005)	-0.010	(-4.200)	1.50%	51
Bank of America	-0.169	(-1.671)	-0.004	(-0.053)	0.501	(7.752)	20.97%	30
Boeing	-0.159	(-1.989)	0.061	(0.701)	0.052	(2.642)	1.42%	50
Cisco	-0.179	(-1.995)	-0.048	(-0.386)	0.039	(0.703)	1.45%	37
Exxon Mobil	-0.410	(-2.636)	-0.231	(-1.377)	3.062	(0.854)	11.67%	39
Facebook	-0.728	(-5.322)	-0.350	(-1.708)	2.347	(6.695)	57.60%	7
General Electric	-0.001	(-0.011)	-0.218	(-2.283)	-0.560	(-1.890)	11.75%	37
Home Depot	-0.303	(-1.822)	-0.387	(-1.363)	0.966	(1.424)	7.05%	31
IBM	-0.209	(-2.208)	-0.202	(-2.051)	0.672	(0.912)	3.57%	68
Johnson & Johnson	-0.177	(-1.156)	-0.418	(-1.791)	1.203	(1.446)	7.52%	27
McDonald's	-0.340	(-2.793)	-0.728	(-5.236)	1.711	(0.498)	22.15%	26
Merck	0.006	(0.053)	-0.360	(-3.467)	0.561	(1.416)	11.56%	46
Metlife	-0.207	(-1.532)	-0.306	(-1.296)	0.384	(0.536)	4.80%	18
Microsoft	-0.044	(-0.579)	-0.208	(-1.652)	0.032	(1.077)	5.56%	52
Monsanto	-0.460	(-4.929)	-0.260	(-1.221)	1.229	(1.368)	16.64%	30
Nike	-0.267	(-2.039)	-0.041	(-0.138)	-0.048	(-1.208)	0.88%	31
Pfizer	0.064	(0.506)	-0.121	(-0.889)	-0.058	(-7.275)	9.54%	26
Procter & Gamble	-0.109	(-0.897)	-0.673	(-2.801)	0.113	(5.212)	13.81%	38
Starbucks	-0.008	(-0.059)	-0.184	(-1.041)	0.147	(0.261)	2.54%	32
Tesla	-0.547	(-1.277)	-0.383	(-1.248)	0.234	(0.127)	8.12%	12
Valero	-0.073	(-0.671)	0.115	(0.554)	-0.001	(-0.023)	0.60%	36
Verizon	-0.083	(-0.827)	-0.166	(-0.632)	-0.258	(-0.131)	1.41%	31
WalMart	-0.384	(-2.401)	-0.243	(-1.090)	2.793	(3.746)	15.27%	35

In order to gain isolated exposure to the risks associated with the two sources of realized variance, namely continuous and discontinuous price movements, I construct two risk-factor-mimicking portfolios from traded options that are intended to offer exposure to one of these risks while being relatively unaffected by the other. Using these two risk factor portfolios, I find evidence that compensation for jump risk in addition to compensation for diffusive risk represents a substantial part of variance swap returns. The contribution of the two types of risk to average variance swap returns appears to be of comparable magnitude.

Even when the Fama and French (1993) factors are included as additional control variables, the coefficients on the two risk-factor-mimicking portfolios remain positive and highly significant, which further supports the view that jump risk and diffusive volatility risk constitute independently priced risk factors that are not already accounted for by commonly used risk factors.

However, none of the applied regression specifications is able to fully explain observed variance swap returns. Regression alphas are negative and significant for virtually all underlyings. This circumstance can either result from inefficient pricing of options and variance swaps that allows to earn substantial abnormal returns which are unrelated to systematic risks, or from exposure to additional risk factors that are not considered in the regressions but priced by the market.

Motivated by the work of Drechsler (2013) who links the variance risk premium to model uncertainty and ambiguity aversion, I consider model uncertainty as one such potential factor and find that higher model uncertainty is associated with a more negative variance risk premium. Even though the chosen specification is relatively parsimonious and potentially neglects several important control variables, these results could therefore point toward a variable that may help explain variance swap returns. An interesting extension for future research could therefore be to find a reliable measure of model uncertainty that is available at a relatively high frequency, for example daily or weekly, and test whether this measure helps to explain variance swap returns.

References

References

Aït-Sahalia, Y. and Lo, A. W. Nonparametric estimation of state-price densities implicit in financial asset prices. *The Journal of Finance*, 53(2): 499–547, 1998.

Aït-Sahalia, Y., Cacho-Diaz, J., and Laeven, R. Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics*, 117(3):585–606, 2015a.

Aït-Sahalia, Y., Karaman, M., and Mancini, L. The term structure of variance swaps and risk premia. *Working Paper*, 2015b.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Ebens, H. The distribution of realized stock return volatility. *Journal of Financial Economics*, 61(1): 43–76, 2001.

Anderson, E. W., Ghysels, E., and Juergens, J. L. The impact of risk and uncertainty on expected returns. *Journal of Financial Economics*, 94(2): 233–263, 2009.

Ang, A. and Chen, J. Asymmetric correlations of equity portfolios. *Journal of Financial Economics*, 63(3):443–494, 2002.

Bajeux-Besnainou, I. and Rochet, J.-C. Dynamic spanning: Are options an appropriate instrument? *Mathematical Finance*, 6(1):1–16, 1996.

Bakshi, G. and Kapadia, N. Delta-hedged gains and the negative market volatility risk premium. *Review of Financial Studies*, 16(2):527–566, 2003.

Bates, D. S. Post-'87 crash fears in the S&P 500 futures option market. *Journal of Econometrics*, 94(1):181–238, 2000.

Bates, D. S. The market for crash risk. *Journal of Economic Dynamics and Control*, 32(7):2291–2321, 2008.

Bekaert, G. and Hoerova, M. The VIX, the variance premium and stock market volatility. *Journal of Econometrics*, 183(2):181–192, 2014.

Black, F. Studies of stock price volatility changes. In *Proceedings of the 1976 meetings of the Business and Economic Statistics Section, American Statistical Association*, pages 177–181, 1976.

Black, F. and Scholes, M. The pricing of options and corporate liabilities. *The Journal of Political Economy*, 81:637–654, 1973.

Bloom, N. The impact of uncertainty shocks. *Econometrica*, 77(3):623–685, 2009.

Bollerslev, T. and Todorov, V. Tails, fears, and risk premia. *The Journal of Finance*, 66(6):2165–2211, 2011.

Bollerslev, T., Gibson, M., and Zhou, H. Dynamic estimation of volatility risk premia and investor risk aversion from option-implied and realized volatilities. *Working Paper*, 2008.

Bollerslev, T., Gibson, M., and Zhou, H. Dynamic estimation of volatility risk premia and investor risk aversion from option-implied and realized volatilities. *Journal of Econometrics*, 160(1):235–245, 2011.

Bollerslev, T., Li, S. Z., and Todorov, V. Roughing up beta: Continuous vs. discontinuous betas, and the cross-section of expected stock returns. *Working Paper*, 2015.

Bondarenko, O. Market price of variance risk and performance of hedge funds. In *Working Paper, University of Illinois at Chicago*, 2004.

Breeden, D. T. and Litzenberger, R. H. Prices of state-contingent claims implicit in option prices. *The Journal of Business*, 51:621–651, 1978.

Britten-Jones, M. and Neuberger, A. Option prices, implied price processes, and stochastic volatility. *The Journal of Finance*, 55(2):839–866, 2000.

Broadie, M., Chernov, M., and Johannes, M. Understanding index option returns. *Review of Financial Studies*, 22(11):4493–4529, 2009.

Cao, H. H. and Ou-Yang, H. Differences of opinion of public information and speculative trading in stocks and options. *Review of Financial Studies*, 22 (1):299–335, 2009.

Cao, J. and Han, B. Cross section of option returns and idiosyncratic stock volatility. *Journal of Financial Economics*, 108(1):231–249, 2013.

Carr, P. and Lee, R. Volatility derivatives. *Annual Review Financial Econonomics*, 1(1):319–339, 2009.

Carr, P. and Madan, D. Towards a theory of volatility trading. *Risk Books on Volatility*, (29):417–427, 1998.

Carr, P. and Wu, L. Variance risk premiums. *Review of Financial Studies*, 22 (3):1311–1341, 2009.

Carr, P. and Wu, L. Leverage effect, volatility feedback, and self-exciting market disruptions. *Bloomberg Portfolio Research Paper*, (2009-03), 2011.

CBOE. The CBOE volatility index - VIX. *Chicago Board Options Exchange, White Paper*, pages 1–23, 2015.

Coval, J. D. and Shumway, T. Expected option returns. *The Journal of Finance*, 56(3):983–1009, 2001.

Cox, J. C. The constant elasticity of variance option pricing model. *The Journal of Portfolio Management*, 23(5):15–17, 1996.

Cremers, M., Halling, M., and Weinbaum, D. Aggregate jump and volatility risk in the cross-section of stock returns. *The Journal of Finance*, 70(2): 577–614, 2015.

Demeterfi, K., Derman, E., Kamal, M., and Zou, J. A guide to volatility and variance swaps. *The Journal of Derivatives*, 6(4):9–32, 1999.

Drechsler, I. Uncertainty, time-varying fear, and asset prices. *The Journal of Finance*, 68(5):1843–1889, 2013.

Drechsler, I. and Yaron, A. What's vol got to do with it. *Review of Financial Studies*, 24(1):1–45, 2011.

Driessen, J., Maenhout, P. J., and Vilkov, G. The price of correlation risk: Evidence from equity options. *The Journal of Finance*, 64(3):1377–1406, 2009.

Du, J. and Kapadia, N. The tail in the volatility index. *Working Paper, University of Massachusetts*, 2011.

Dupire, B. Pricing with a smile. *Risk*, 7(1):18–20, 1994.

Ellsberg, D. Risk, ambiguity, and the savage axioms. *The Quarterly Journal of Economics*, 75:643–669, 1961.

Engle, R. and Figlewski, S. Modeling the dynamics of correlations among implied volatilities. *Review of Finance*, 2014.

Fama, E. F. and French, K. R. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56, 1993.

Fama, E. F. and French, K. R. The capital asset pricing model: Theory and evidence. *The Journal of Economic Perspectives*, 18:25–46, 2004.

Fournier, M. and Jacobs, K. Inventory risk, market-maker wealth, and the variance risk premium: Theory and evidence. *Rotman School of Management Working Paper*, (2334842), 2015.

Gârleanu, N., Pedersen, L. H., and Poteshman, A. M. Demand-based option pricing. *Review of Financial Studies*, 22(10):4259–4299, 2009.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801, 1993.

Gourier, E. Pricing of idiosyncratic equity and variance risks. *Working Paper Princeton University*, 2015.

Gourio, F. Disaster risk and business cycles. *American Economic Review*, 102: 2734–2766, 2012.

Green, T. C. and Figlewski, S. Market risk and model risk for a financial institution writing options. *The Journal of Finance*, 54(4):1465–1499, 1999.

Herskovic, B., Kelly, B., and Lustig, H. The common factor in idiosyncratic volatility: Quantitative asset pricing implications. *National Bureau of Economic Research Working Paper 20076*, 2014.

Heston, S. L. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6 (2):327–343, 1993.

Hull, J. *Options, futures and other derivatives*. Pearson/Prentice Hall, 2009.

Hull, J. and White, A. The pricing of options on assets with stochastic volatilities. *The Journal of Finance*, 42(2):281–300, 1987.

Hull, J. and White, A. OIS discounting, interest rate derivatives, and the modeling of stochastic interest rate spreads. *Journal of Investment Management*, 13:64–83, 2015.

Ilmanen, A. Do financial markets reward buying or selling insurance and lottery tickets? *Financial Analysts Journal*, 68(5):26–36, 2012.

Jiang, G. J. and Tian, Y. S. The model-free implied volatility and its information content. *Review of Financial Studies*, 18(4):1305–1342, 2005.

Jiang, G. J. and Tian, Y. S. Extracting model-free volatility from option prices: An examination of the VIX index. *The Journal of Derivatives*, 14(3):35–60, 2007.

Kelly, B. and Jiang, H. Tail risk and asset prices. *Review of Financial Studies*, 27(10):2841–2871, 2014.

Lee, S. S. and Mykland, P. A. Jumps in financial markets: A new nonparametric test and jump dynamics. *Review of Financial Studies*, 21(6):2535–2563, 2008.

Merton, R. C. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3(1-2):125–144, 1976.

Neuberger, A. The log contract. *The Journal of Portfolio Management*, 20: 74–80, 1994.

Newey, W. K. and West, K. D. A simple, positive semi-definite, heteroskedas-

ticity and autocorrelation consistent covariance matrix. *Econometrica*, 55: 703–708, 1987.

Pan, J. The jump-risk premia implicit in options: Evidence from an integrated time-series study. *Journal of Financial Economics*, 63(1):3–50, 2002.

Ross, S. A. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341–360, 1976.

Schürhoff, N. and Ziegler, A. Variance risk, financial intermediation, and the cross-section of expected option returns. *CEPR Discussion Paper No. 8268*, 2011.

Staum, J. Incomplete markets. In Birge, J. R. and Linetsky, V., editors, *Handbooks in operations research and management science*, chapter 12. Elsevier, 2007.

Todorov, V. Variance risk-premium dynamics: The role of jumps. *Review of Financial Studies*, 23(1):345–383, 2010.

Wu, G. The determinants of asymmetric volatility. *Review of Financial Studies*, 14(3):837–859, 2001.

Junior Management Science 1 (2016) 1-300

Junior Management Science

journal homepage: www.jums.academy

A Common Ground in Guerilla Marketing – State of Research and Further Research Opportunities

Lennart Wendland
Otto-von-Guericke-Universität Magdeburg

Abstract

Guerilla Marketing is a cloudy concept. While practitioners and scientists increasingly use it in their practice, its definition remains not straightforward. This thesis builds a common ground in Guerilla Marketing through clearing up the field. Guerilla Marketing is defined and an overall model is presented that integrates previous efforts. This model has some advantages. Most strikingly, it introduces two highly relevant mediators which include recipient's behavior into the model.

Keywords: Guerilla Marketing, Marketing science, Conceptual framework, Philosophy of science

1. Introduction

"Guerilla Marketing is a body of unconventional ways of pursuing conventional goals" is how Levinson (2008) broadly describes a concept which he himself devised and who claimed to be the "father of Guerilla Marketing" (Schulte, 2007, p. 16). This ambiguous definition was obviously only one of Levinson's innumerable contributions since the early 1980s, yet it reflects that the concept itself leaves plenty of room for interpretation. Over the past thirty years practitioners and scholars have used the term for a wide range of activities, instruments, concepts, tools, strategies and methods, only agreeing on one simple thing: Guerilla Marketing always aims at achieving maximum effects at low expenses (Baltes and Leibing, 2008). Within these parameters it seems that everything else around the concept is left up to the individual who subjectively defines it for whatever reason desired. A diffuse understanding, lacking clear-cut definitions and practitioners who constantly (re-)invent "unconventional" marketing concepts under the umbrella of Guerilla Marketing are the result of this confusion. Guerilla Marketing and whatever action one derives from it has become the random synonym for almost any marketing or advertising activity that does not fit the classical frame. Furthermore, this misapprehension carries on from practice to scholarship as "literature still lacks a scientific contribution which directly focuses on the guerilla concept" (Hutter and Hoffmann, 2011a, p. 2).

To recap, a mass of activities and concepts can be found that are said to be aligned with the Guerilla Marketing concept but only little input has been delivered on how to define, classify or categorize those approaches. A final scientific model to explain the entire concept is still missing and so far mostly single actions or effects have been described. The definition of Levinson (2008) as a basis is just as explicit as it is misleading yet it still today is of a tremendous relevance for a critical discussion on the topic. Asking whether or not Guerilla Marketing really is as innovative as it is said to be or if it simply can be perceived as a trend that managed to survive evolution as a subculture to marketing in general needs further questioning. Due to its relevance it has been carried on over the decades and still causes confusion nowadays.

1.1. Heritage

Ernesto Che Guevara Lynch de la Serna, the leader of the Cuban revolution, delineated the Guerilla tactics as a method to campaign war through surprising ambush attacks allowing a practically inferior army to succeed over the outnumbering opponent (Guevara, 1982). The inferior Guerilla warriors avoided the open battle and rather made use of surprise effects and acts of sabotage versus the military and even their own government (Schulte and Pradel, 2006). The term itself linguistically derived from the Spanish word "guerra" simply meaning "war" (Nufer and Bender, 2008). The adjustment "Guerilla" therefore can be translated to "lit-tle war" (Puttenat, 2007).

DOI: http://dx.doi.org/10.5282/jums/v1i1pp34-59

Transferring these tactics to marketing, practitioners have quickly adapted and created the term "Guerilla Marketing" during the mid-1960s (Baltes and Leibing, 2008). During that time a change from a seller- to a buyer-market took place and companies felt the need of creative ideas, ingenuity and flexibility in their effort to persuade the consumers of their products and brands (Schulte and Pradel, 2006). In its early days mostly small and middle scale enterprises (SME's), with strongly limited budgets, used these revolutionary strategies trying to get the attention of the consumer and moreover doing whatever it takes to weaken competition (Puttenat, 2007). Over the years, as external factors changed, the range of companies trying to benefit from Guerilla Marketing actions strongly increased independently from the original reasons. The effectiveness of traditional forms of advertisements constantly decreased (Smith et al., 2007). New pressures on companies and consumers developed as customers started to show lower brand loyalty and greater eagerness to switch between competing brands (Roy and Chattopadhyay, 2010). Word-of-mouth (WoM) gained in relevance as mistrust in company's messages made classical channels defective and companies incapable of distributing and diffusing their content into society (Keller and Berry, 2003). Most recent incidents like the financial crisis of 2008/09 are good examples of influential factors facilitating the demand for cheap alternatives as marketing budgets are quickly and easily reduced in rough times. Factors like those also led the larger players in the business to rethink their strategies focusing on more cost-effective methods to differentiate (Porter, 1985).

While in its early days the concept was mostly appealing to SME's competing against the big players in fiercely competitive markets, Kotler et al. (2007, p. 12) states that "Guerilla warfare is normally practiced by smaller companies against larger companies" deserves critique; even the Goliaths of the industries nowadays take advantage of this method which has the potential to reach a great amount of customers, cheaply and cost-efficiently, while seeking an undisturbed dialogue with the consumer maximizing the impression left behind (Kotler et al., 2007; Bigat, 2012). For those big companies, entering new territory obviously comes along with a certain risk, yet despite the budgetary pressure, the changing environment and while having to target market segments as heterogeneous towards others as possible, it might be worth taking such risk (Ansoff, 1965). Since the 1980s Jay Conrad Levinson, strongly promoted the Guerilla Marketing evolution through his various contributions and pushed the concept to its final breakthrough. Ever since marketers and scholars have used this term referring to any kind of consumer communication instrument that aims at optimizing the cost-benefit ratio through "unconservative" means (Hutter and Hoffmann, 2011a).

Over the years marketing literature has identified more and more concepts and instruments of unconventional advertisement techniques (e.g. Buzz Marketing, Viral Marketing, Ambient Marketing, Ambush Marketing, Celebrity Marketing) whereas many further randomly categorized further as Guerilla Marketing or at least used in close relation by practitioners in everyday life (Mughari, 2011). Yet, great descriptive and normative confusion about what scholars and practitioners really refer to when labelling an instrument or campaign "unconventional" can be witnessed. With many deficits already identified, the extensive relevance for the topic will be presented in the adjacent chapter.

1.2. Relevance

The necessity for critically reviewing and the relevance of this topic in general from a practical perspective can be found in the exemplary study of the GfK (2009) . Matching data from the years 2005 and 2009 the association's study compared the use of conventional and (professed) Guerilla Marketing methods and instruments through interviewing 233 marketing associates and managers (GfK, 2009). As a result the study showed that the use of television advertising had decreased by 1.5% and radio advertising by 5.3% whereas contrarily the use of Ambient Marketing had increased by 5.4% and Viral marketing by even 12.8% (GfK, 2009).

Apart from a practical perspective describing increased interest in the operational field, the *"shortage , up to now, of scientifically substantiated knowledge on the subject matter of guerilla marketing, its instruments and its categorization may be interpreted from two different points of view: guerilla marketing cannot be classified or guerilla marketing is difficult to classify"* (Nufer, 2013, p. 5).

While searching for relevant literature one quickly realizes how little developed scientific research is so far and that research is lost in a theoretical ambiguity. The existing definitions deliver a basic understanding but just now started to present all-embracing projections and a lacking terminological delimitation results in an imbroglio in literature (Zerr, 2003). The lack of unified and especially overreaching definitions does not allow scholars to speak the same language and therefore gain a deeper insight into the concept. Still definitions are facilitated by strategical orientations and scholars interpreting the concept from their own specialized perspective (Tropp, 2011). Various practitioners have distributed articles bearing only specialized scientific results, upholding examples of executed actions. In fact many studies are mainly driven by practitioners and scholars trying to explain or justify single advertising actions or examples instead of developing scientific models as a basis for exploration (Yüksekbilgili, 2014). They are primarily interested in whether or not the single effect they desired is really achieved. An interplay of effects as a basis to derive and test models has not been sufficiently delivered. Furthermore parameters and metrics to actually measure effectiveness and delineate accordingly are missing. Exemplary approaches measure the effectiveness of the ratio of output and costs whereas output does not refer to final sales but rather number of recipients exposed (Hutter and Hoffmann, 2013).

Overall scientific critique is lacking. The authors barely comment on which classifications need to be fulfilled in the development of Guerilla Marketing conceptions. As a consequence to lacking models, methodology for Guerilla Marketing and its concepts is rarely presented as a whole. Science

needs to catch up to the trend of such a dimension. Only few empirical studies which would try to measure the complete effectiveness of Guerilla Marketing exist while those few aiming at explaining the mechanisms of the concept need revision (Hutter and Hoffmann, 2014). The existing models trying to explain the concept or measure the efficiency need to be aligned and carried forward. In those models presented only partial understanding can be drawn from given studies as they primarily operationalize single mediator and moderator variables of a concept that as a whole still lacks an entry. Even those novel distributions manifest deficits as the initial position is often left unclear whereas exemplary influential factors like behaviorism and deeper insights into the psychological processes are being ignored or only briefly and fragmentarily addressed (Hutter and Hoffmann, 2013).

As instruments and tools that used to be innovative and different from classical advertising have started to become part of the classical marketing world, one could expect this evolution to continue in the future. What is perceived as a breathtaking reformation in practice today could, and probably will, become routine in the future; or already has become ordinary today. Assimilation is therefore vital for science to be up-to-date. The effect of a maturing approach has mostly been disregarded so far apart from assigning a higher relevance.

However, to ask how unconventional the Guerilla Marketing concept really is and to what extent nowadays one can already refer to as a mature advertising strategy that has simply never gained the attention in science that it deserves and requires, is left unanswered.

1.3. Research question

In the process of accomplishing a common ground in Guerilla Marketing through clearing up the field and with the goal to then deliver propositions for further research several key questions will need to be considered. As most of the current problems have already been addressed beforehand, the upcoming short reasoning will justify the necessity for the research questions identified. Derived from the relevance inherent in the issues described above, obviously there is a call for action.

At first classifications are vital for further approaches and deeper insights. Considering the field of marketing in general, the position of Guerilla Marketing within and the ambiguity arising from the misleading use of terminology will be supportive. The following question needs to be considered on the path to categorizing Guerilla Marketing:

> I) How can Guerilla Marketing be arranged in the frame of reference?

Furthermore circumventing Guerilla Marketing from conceptual features and classifying the concept will be necessary when assigning exclusive characteristics on the way to an all-embracing definition.

> II) How can Guerilla Marketing be circumvented?

> III) How can Guerilla Marketing be defined regarding the conceptual classification and delimitation?

Apart from identifying an up-to-date and comprehensive definition, scientific models shall illustrate the complex relationships of the Guerilla Marketing concept. Hence it is highly relevant to ask:

> IV) How can Guerilla Marketing be modelled?

In addition the presented models need testing and a scientific empirical foundation. Subsequently the models shall further serve as a frame for presenting relevant scientific studies.

> V) How can Guerilla Marketing models be empirically founded?

The following course of action is supported by the use of major Roman lettering for addressed research questions which will already be answered and briefly discussed within the according résumé and critique in chapters themselves. A further listing can be found in the conclusion of the thesis with direct reference to the roman lettering whilst the propositions and opportunities for further research will be covered in a separate chapter at the end of this thesis.

1.4. Structure and methodology

On the way to a common ground in Guerilla Marketing this thesis presents a total of 8 chapters which will cover the most relevant factors of the concept subsuming with propositions for further research opportunities. Throughout chapters 2 to 6 a résumé and a critique will conclude each chapter. The résumés serve as a brief recall of the presented information from the according chapters whereas the critiques entail the author's critical reflection of the content including own contributions.

The introduction in chapter one presents a first glance at the field including germane information regarding the heritage and necessity of the practical field of Guerilla Marketing and further delivers a justification for the relevance of scientifically discussing the topic. The research questions in chapter 1.3 set the framework and serve as a guide through the thesis. Chapter 2 fits Guerilla Marketing into the field of marketing in general and further presents how the concept is aligned in the marketing mix. Classifying Guerilla Marketing through ambiguously interpreted "unconventionality" will be discussed alongside.

Subsequently and with given classification approaches in mind the path towards defining Guerilla Marketing will be covered in chapter 3. Here the strategical orientation of the respective scholar builds the basis for classifying the most relevant definitions.

Due to its practical nature the most acknowledged Guerilla Marketing instruments will be presented in chapter 4. Examples will be presented in the appendices. Modelling Guerilla marketing and critically discussing the latest contributions will again be concluded with suggestions of the author in chapter 5. The author's model as an extension of the

current state of research will serve for further discussing the empirical research presented regarding Guerilla Marketing. Testing and operationalizing through empirical research will be vital for further scientific progress and therefore leads towards the presentation of further research opportunities and a final discussion in chapters 7 and 8.

This thesis will present results exclusively on the basis of literature research. Only secondary data is being used. As described above, the according research questions will be derived within the chapters, summarized at the end of them and presented and discussed at the end of the thesis in order to fully revise. The final discussion will briefly gather and review the results and subsequently propositions for further research will be presented.

2. Marketing

Linguistically Guerilla Marketing can easily be identified to fall under the umbrella of the field of marketing in general. Yet after having presented the heritage of the concept as well as the practical and scientific relevance, closer attention needs to be drawn towards finding Guerilla Marketing within the field, especially the marketing mix, to then be able to further classify, delimitate, frame and also define Guerilla Marketing in succession.

2.1. Guerilla Marketing in the field of marketing

Marketing is a discipline which tries to influence the needs and the demands of individuals through making products attractive and available, always upholding the premise of satisfying individual's needs instead of actually creating them (Kotler et al., 2007). Therefore nowadays every action an organization takes has to be aligned with the needs of the consumers, constantly bearing in mind that the relevant criteria for making a purchase decision are always the consumer's subjective perceptions of the totality of a product or service.

"Marketing is an entrepreneurial way of thinking (. . .) concretized in the analysis, planning, execution and control of any internal and external corporate action, which, through the orientation of a company's outputs towards the customer's needs (. . .), aims at achieving an organization's outlet markets' goals" (Bruhn, 2007, p. 14).

In his definition Bruhn describes marketing as a key concept of the management theory. He further postulates that if an organization manages to increase the perceived value of a good, then the organization was able to create additional utility as a strategical competitive advantage (Bruhn, 2007). Interestingly, Bruhn (2007) also states that successful decision making and problem solving in marketing theory, apart from analytical approaches, highly depends on innovative and creative ideas. Consequently, the target of any marketing effort is to positively influence the individual's perception through the use of the marketing mix, especially consumer communication (Böcker and Helm, 2003).

Historically evolution of marketing-orientation developed drastically during the past 50 years. Having left behind the strong supplier and sale oriented phase of the 1970s, marketing-orientation shifted towards the consumer's perspective (Hutter and Hoffmann, 2013). Between the 1980s and 1990s a further shift towards a competition-orientation was discovered while since the 1990s a strong relationship-orientation was paramount (Hutter and Hoffmann, 2013). Nowadays, building relationships with consumers is still of a very high relevance whereas a shift towards a value-oriented marketing is taking place. Focusing on strong relationships and efficiently using bindings to profitable and valuable consumers (e.g. through engaging in positive WoM) can currently be witnessed (Hutter and Hoffmann, 2013). The importance and efficacy of WoM is being rediscovered primarily through modern technologies.

The evolution of marketing practice has changed dramatically within the twentieth century. Technological progress and the introduction of printing technology allowed a shift towards a much more creative execution of marketing advertising to occur and further made it possible to suddenly reach out to a much larger group of consumers (Egan, 2007). The introduction of the radio and later television complemented and enhanced the shift whereas both, later on, became the predominant advertising vehicle followed by internet (Egan, 2007).

Today marketing plays a major role in our everyday life, confronting an individual with up to 1.500 advertising messages daily meaning that the proportion of content in which organizations try to communicate with desired recipients makes up over a quarter of those six hours an individual consumes those media (Zerr, 2003).

2.2. Marketing mix

The marketing mix is the sum of all tools an organization can use in combination to provoke the desired reactions on their target markets (Meffert et al., 2011). Those tools are traditionally represented by the "4Ps", referring to *Product, Price, Place* and *Promotion*, whereas additional tools have been identified in various other models (Kotler et al., 2007). As the presentation of the marketing mix will serve a later classification of Guerilla Marketing, those further contributions will not be considered. The art of combining those traditional 4Ps is vital for achieving marketing strategies and goals whereas all four tools are strongly interdependent and cannot be executed independently (Meffert et al., 2011).

Product constitutes all decisions an organization makes towards creating a product which satisfies consumers' needs and adds additional value (e.g. naming, branding, packaging, innovation and assortment) towards achieving a competitive advantage (Bruhn, 2007). *Price* refers to any decision that is being made towards the actual sales-price of a product but also includes additional amendments like discounts, contracts regarding terms and conditions and payment and delivery options (Bruhn, 2007). *Place* transcribes any action and decision that handles the supply of the consumer with an organization's products focusing on sales, logistics and also intermediaries like wholesalers (Bruhn, 2007).

Promotion includes the presentation to and communication with the target groups defined through any given instrument which is being used to interact with the designated recipient (Bruhn, 2007). Hence the communication shall also result in increased sales (Kotler et al., 2007). The five fundamental instruments of Promotion are advertising, sales promotion, public relations, direct sale and direct marketing (Kotler et al., 2007). Those could further be classified into internal, interactive and external communication, whereas external communication, falling under this category will be discussed, as a matter of relevance, in further detail (Bruhn, 2007).

It is highly relevant to notice that concepts like Guerilla Marketing have been acknowledged to potentially be applicable to all four tools of the marketing mix (Schulte and Pradel, 2006; Kotler et al., 2007). Guerilla Marketing does not compete with other marketing strategies of the marketing mix but rather complements the original tools. Yet, Guerilla Marketing has major influence on and also finds its primary area of application within the communication region. Schulte and Pradel (2006) presented a study giving proof that approximately 70As subsequently described in detail, the classical marketing mix and the way advertising is represented within shall be understood as a classical component and in succession will be used in discussions as contraire to non-classical marketing and advertising approaches (Bruhn, 2007). Hence, Guerilla Marketing and its instruments can adaptively be very well integrated into classical conceptual marketing frameworks as part of the operational marketing planning and execution (Meffert et al., 2011).

2.3. Classification through unconventionality

Due to the fact that ambiguity causes most issues, the Guerilla Marketing concept and its instruments will be presented in alignment with the term "Marketing" and "Advertising" interchangeably. In later sections the topic of abstract concepts will still be highlighted and discussed. Exemplary and anticipatory "Ambient Marketing" or "Ambient Media" are ambiguously used in literature and practice as well and both carry various synonyms. In the further reading Guerilla Marketing and its concepts are therefore titled with the "Marketing" appendant, instead of using intuitive and also commonly used appendants like "Media".

Further propositions of aligning the terminology (e.g. changing the title from Guerilla Marketing to "Mac-Gyver Marketing") have frequently been proposed yet gained only moderate attention while the content behind the umbrella term is still unclear (Patalas, 2006). Apart from titles the delineation in regard to what is considered "unconventional", "classical" or "innovative" as a base for most definitions will be presented in detail in the following section.

To circumvent the Guerilla Marketing concept one at first needs to understand what "conventional" or "unconventional" advertising really refers to. Here, in reference to Guerilla Marketing, the terms "unconventional", "alternative" and "non-conservative" will be, just as "conventional", "classical" and conservative", used as interchangeable synonyms. Those scholars assigning these characteristics to Guerilla Marketing use the terms comparably and refer to almost identical initial positions (Hutter and Hoffmann, 2011b; Schulte and Pradel, 2006; Nufer and Bender, 2008; Zerr, 2003; Patalas, 2006; Levinson, 2008).

Due to the fact that classical advertising is communicated via mass-communication media, Bruhn uses the term "media-advertising" to describe classical forms of advertising in newspapers and on boards or in television- and radio spots (Bruhn, 2007, p. 204). These forms actually refer to the spot or space that is used within a medium. The medium in which the message is distributed in is referred to as an "advertising-vehicle" (Bruhn, 2007). A typical vehicle would be a magazine, newspaper or television whose publishers, operators or broadcast stations allow for external organizations to book space and time for the purpose of communicating foreign content. The main characteristic of these forms is the unilateral way of communicating. The recipient has no chance of interacting with the sender in any way (Bruhn, 2007). Hence, Bruhn states that it is the transmission and distribution of promotional content via booking "media-advertising" in advertising-vehicles" in public spaces with payments made in return (Bruhn, 2007, p. 204). Furthermore, here the internet is considered a mass-communication medium, and therefore "media-advertising", which on the one hand, can be used as a classical advertising vehicle but, on the other hand, bears potential to serve for unconventional advertising approaches while contradicting the classification unilateral form of communication (Bruhn, 2007). Due to the mass of opportunities and its multifaceted scope of application, the internet as a mass-communication medium has a special standing. Apart from a distinction between mass-media communication, a classification of advertising measures can be found in the distinction between "above-the-line" and "below-the-line" (Nufer and Bender, 2008)(see also Figure 1). Contrary to the situation in the middle of the twentieth century, by the year 2000, two-thirds of the amount of money spent on marketing communications was spent "below-the-line" (Varey, 2002). Below-the-line, as defined by Egan (2007) simply stands for any marketing instrument that is not advertising in the classical sense. Contrarily and in this content referred to from a classical sense, "above-the-line" refers to anything the recipient can immediately identify: posters, advertising boards, television-spots or radio advertising (Egan, 2007, p. 11). Examples for below-the-line methods are public relations, direct marketing or event marketing (Nufer and Bender, 2008; Gelbrich et al., 2008). From another perspective above-the-line mostly corresponds to what Bruhn (2007, p. 204) has defined as "media-advertising" which can be "booked" for a "performance-related payment".

2.4. Résumé

Finding Guerilla Marketing and arranging it as an all-embracing concept which can be applied to any marketing tool is quite comprehensible. Hence classifying the Guerilla

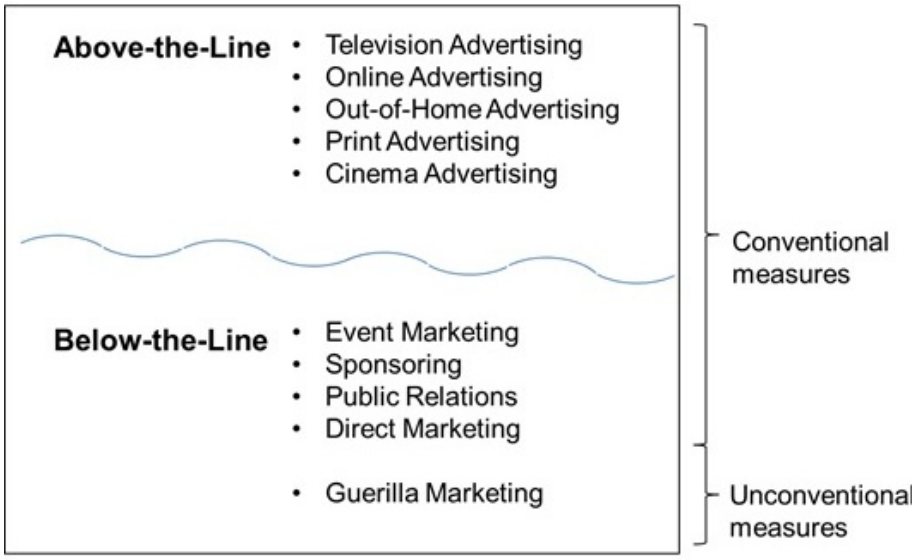

Figure 1: Guerilla Marketing in communication politics; Source: following Gelbrich et al. (2008, p. 178).

Marketing concept as part of the marketing mix due to its common operational use seems intuitive and reasonable.

Yet, delimitating Guerilla Marketing in a way that assigns exclusive features appears to be the struggle that has not yet come to a conclusion. The discussions and later definitions deliver ambiguous conclusions in regard to the categories of "above-the-line" and "below-the-line", "media-advertising" or not and most relevantly "unconventional" and "conventional" and will deserve closer attention in the following critique (Hutter and Hoffmann, 2013; Bruhn, 2007; Nufer and Bender, 2008).

2.5. Critique

As a result to continuous discussions, it has to be noticed that the distinction between "above" or "below-the-line" and "(un-)conventional" does not imply equal constraints on Guerilla Marketing. While Egan (2007, p. 11) states that "below-the-line" refers to any tool which is "not advertising in the classical sense", this statement does not refer to a distinction between "conventional" and "unconventional" (Gelbrich et al., 2008). As the classification between "above"- and "below-the-line" is perceived as rather "spongy" and needs further containment, for now the distinction between what scholars and practitioners perceive as "conventional" or "unconventional" is of a rather bigger focus (Hutter and Hoffmann, 2013, p. 11). Gelbrich et al. (2008) depict Guerilla Marketing, among others, as a "below-the-line" tool or instrument. Still, it is the only one that is titled as "unconventional" bearing plenty of space for interpretation. As a concept that has been identified to potentially work within every tool of the marketing mix unbound to vehicles one first needs to revise and consider the fields of application (Schulte and Pradel, 2006). An own depiction following prior contributions is proposed by the author (Gelbrich et al., 2008; Hutter and Hoffmann, 2013).

In the context of Guerilla Marketing and disregarding its establishment, the critique on an above- and below-the-line classification is comprehensible. Just through considering the characteristics of Ambient Marketing presented in chapter 4.1.4 will strongly support objections towards the application. The example clearly shows how an "above-the-line" method can easily move "below-the-line" of visibility for the consumer.

Hence, the classification of Guerilla Marketing needs further attention. Following the definition of media-advertising by Bruhn (2007) and assigning every other method to be "unconventional", then, following Gelbrich et al. (2008) and Hutter and Hoffmann (2013), as a consequence "unconventional" only and exclusively refers to Guerilla Marketing "below-the-line" (Bruhn, 2007; Hutter and Hoffmann, 2013).

Yet, in this regard and as depicted in figure 2 Guerilla Marketing should at least be considered from two perspectives: either as a method that clearly circumvents from classical media-advertising frames (e.g. anything that can or cannot be booked in an advertising vehicle) or as a principle (e.g. philosophy) that could potentially be applied to practically any method or tool available (Bruhn, 2007; Levinson, 2008). Such bidirectional perspective would enhance further approaches yet today a classification presented above is still predominant. "Unconventionality" could therefore be replaced or at least as a term become more specific with complementing explanations.

3. Guerilla Marketing

After having addressed the need for, the heritage of and a first conceptual categorization within the field of marketing, further chapters will deliver deeper insights. Still today no uniform acknowledged definition for Guerilla Marketing has been established. It is still not entirely clear, how Guerilla Marketing can be identified or described as such or which

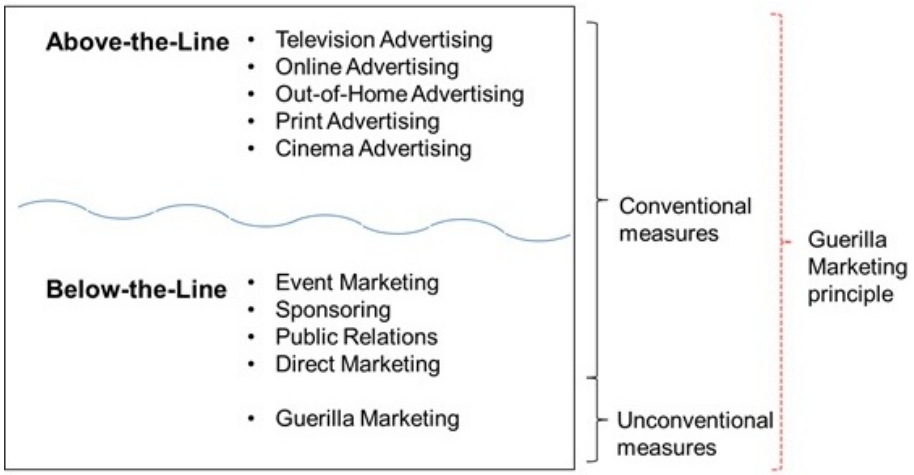

Figure 2: Guerilla Marketing in communication politics; Source: following Gelbrich et al. (2008, p. 178)

instruments really fall under the umbrella of Guerilla Marketing (Patalas, 2006). Problems arise mainly from lacking terminological distinction, a change of perspective over time and an orientation towards the content. Especially the terminological confusions are further being enhanced by practitioners who on the one hand constantly (re-)invent terms which in the end all can be classified as part of or at least very closely related to the Guerilla Marketing concept and on the other hand lacking scientific progress in delivering a clear partition (Zerr, 2003).

Hence, through the first approaches, the discussion on terminology and especially applicability, focus points for the upcoming sections have been made visible in order to derive the presented definitions and critically reflect in order to develop a common ground in Guerilla Marketing.

3.1. Defining Guerilla Marketing

The strategical orientation of any given definition most suitably identifies the perception of the author. This orientation could potentially be competition-oriented or consumer-oriented (Hutter and Hoffmann, 2013).

Apart, scientists, in their definitions, nowadays start to tackle the Guerilla Marketing concept as whole and have quit trying to put their focus on only single instruments (Hutter and Hoffmann, 2013). This development is said to be caused by the fact that Guerilla Marketing as a concept has matured and therefore gained increased attention in practice and science. While when the basic concept had first been adapted during the 1960s to 1980s it mainly contained the idea of achieving a maximum output with only little input, during the phase of differentiation in the life cycle of Guerilla Marketing the concept primarily fed on creative, flexible and surprising ideas (Hutter and Hoffmann, 2013). While some of the earlier definitions also proposed a restrictive characteristic of Guerilla Marketing in only or mainly being suitable for SME's, this allocation is mostly considered to be outdated (Levinson, 1998; Kotler, 2002; Ries and Trout, 1986). Interestingly primarily American contributions still align with this point of view while European scholars follow the view of Guerilla Marketing as an umbrella term for innovative advertising instruments (Krieger, 2012). Ever since the late 2000s and while more and more larger organizations took notice and made use of Guerilla Marketing, trends are now emerging towards regarding Guerilla Marketing as a conception and superordinate marketing strategy and organizations therefore align their entire marketing mix along and with the concept (Hutter, 2013). Exemplary a fundamental marketing conception can easily be adapted to Guerilla Marketing as primarily operative and executive factors determine the adjustments (Meffert et al., 2011).

Tropp (2011) proposed the focus on four basic strategical orientations of Guerilla Marketing definitions: the "David vs. Goliath", the "Assault", the "Communication" and the "All-embracing" strategy aligned with a chronological paradigm shift. The definitions in the upcoming section will therefore be attached to those strategical orientations with focus on the later developments and current contributions.

3.1.1. David vs. Goliath strategy

Levinson (1998) rarely tried to actually deliver definitions of Guerilla Marketing but rather contrasted it with the standard methods upholding the strategical orientation of SME's trying to fight larger opponents. Expansive contributions like the quote opening the introduction to this paper were the origin (Levinson, 2008). Historically the arguments lay in the strong developments of marketing during this period. In his opinion Guerilla Marketing really is about being able to discover every given opportunity and make use of every aspect of marketing (Levinson, 1998). It is said to be an alternative and not a continuance of the classical marketing. Those companies, which are able to afford standard methods, will usually make use of them whereas smaller, younger and less wealthy companies will have to rely on their creativity and finding their potential in Guerilla Marketing activities (Levinson, 1998). Furthermore those smaller companies profit from a time-benefit when making decisions and

therefore are able to act more flexibly and time sufficiently (Levinson and Godin, 1996). Levinson, in his later publications (e.g. Levinson and Godin 1996), revised his view that only small companies can benefit from the Guerilla concept but in fact one can derive where the original concept came from (Levinson and Godin, 1996). For Levinson (2008) successful Guerilla Marketing can only succeed if a real-life story is being told while the consumer's problems are being solved through the product or service. He sees Guerilla Marketing rather as a philosophy or an attitude.

3.1.2. Assault strategy

In one of the early contributions in 1986 Jack Trout and Al Ries defined Guerilla Marketing as a constructive marketing strategy used in the battle of small and medium sized businesses (Ries and Trout, 1986). The authors identified three main principles determining successful Guerilla Marketing. First, a market segment has to be found that is small enough to defend it. It has to be a specific niche that only desires limited resources and that can potentially be determined for example through geographical, product characteristics, a specific target group or branch. Secondly the administrative structure of an organization engaging in successful Guerilla Marketing has to stay slim in order to save costs and quickly react to external changes. Lastly, also determined by the factor of time, an organization has to maintain high flexibility in order to quickly retreat from less attractive and profitable fields (Ries and Trout, 1986).

Not surprisingly, since Jack Trout is one of the pioneers of the positioning strategy in marketing, his efforts primary focus points lay in the strategical targeting and positioning of and in market segments. The authors subsequently focused on instruments like Mosquito Marketing, an instrument of the Guerilla Marketing concept (see chapter 4), and therefore delivered a highly restrictive definition mostly focusing only on a single approach which primarily seeks assaulting competitors.

3.1.3. Communication strategy

"Guerilla Marketing is a contemporary interpretation of a strongly customer-oriented Marketing with a predominant focus on communication politics" is one of the contributions by Patalas (2006) stressing the relevance of, without limiting the field of application to, communication politics in the marketing mix.

Guerilla Marketing will not replace conventional marketing but rather complement it. It's a creative and modern method to break through classical marketing boundaries and necessary whenever the consumer is no longer willing to react to classical marketing communication. In this case it is of a high relevance to invent and make use of new creative methods (Patalas, 2006). The message, the corporate identity and the story of any action needs to match the targeted consumer in an understandable and credible way (Patalas, 2006). Furthermore Patalas proposes the use of the term "Mac-Gyver Marketing" rather than Guerilla Marketing as it uses the given potential of an individual situation or organization and excels through a surprising and quick appearance and disappearance. For Patalas the size of an organization again plays a crucial role as he states that Guerilla Marketing actions usually work on two levels: the "first publicity" is those people who are confronted with an action directly and the "second publicity" is those people that are being reached through word-of-mouth or media broadcastings (Patalas, 2006). The first publicity shall act as a multiplier and the second publicity is the even harder to target mass of recipients which, as a target group, is obviously most interesting to larger companies rather than small and middle scale enterprise (Patalas, 2006). Patalas (2006) therefore delivered a definition that does stress a focus in Guerilla Marketing on SME's but rather states that the size determines the goals along Guerilla Marketing actions. Yet, from his perspective, the strategical focus still mainly lays on consumer politics rather than an integrated approach. In the end it is to ask whether the organization wants to use Guerilla Marketing for generating a strong impression or effect or if it measures the effect through the amount of recipients reached in total (Patalas, 2006).

Schulte and Pradel (2006) depicted a list of constitutive characteristics which were identified as contributions to most of the (then) current definitions. They summarize the postulates stating that Guerilla Marketing bears the following properties: "unconventional, surprising, creative, provoking, cost-efficient, flexible, extraordinary, funny, spectacular and infectious" (Schulte and Pradel, 2006, p. 22). The authors themselves further state that Guerilla Marketing does in fact not depend on the actual operation or execution of an activity, but rather the spectacular and creative idea that is then being executed in a startling action.

"Guerilla Marketing is the art of eliciting the highest level of attention from (. . .) the consumer through unconventional, respectively original, marketing. Thereto it is necessary that the Guerilla Marketeer preferably (but not necessarily) acts outside classical advertising channels and marketing traditions" (Schulte and Pradel, 2006, p. 22).

Guerilla Marketing is generally described as unconventional, innovative, inventive and cost-efficient (Schulte and Pradel, 2006). The authors primarily stress that Guerilla Marketing works through distinction from competition through creative and innovative actions. They further perceive the non-classical form of communicating with the consumer as a strategical orientation rather than an operational execution and therefore deliver a definition with space for the creativity of the executer (Schulte and Pradel, 2006). Schulte himself also further understands Guerilla Marketing as a philosophy and assigns to it any sort of innovative and modern marketing communication tool always upholding the consumer perspective (Schulte, 2007). Krieger (2012) agreed in most points yet included the measure of WoM as a medium to diffuse content and further stressed the perspective of the target group. Only if the target group perceives the communication as unconventional, it has the potential of being carried on (Krieger, 2012). He states that *"Guerilla*

Marketing (in the narrow sense) describes alternative marketing strategies and instruments in out-of-home advertising, whose primary goal it is to obtain great attention with the target groups and stimulate WoM through (out of the view of the target group) unconventional means of communicating outside the classical and established communication channels" (Krieger, 2012, p. 14).

Apart from focusing on target groups Krieger further describes a limited zone of application for Guerilla Marketing by restricting the instruments to out-of-home advertising. His definition is moving closer towards an all-embracing approach despite the out-of-home emphasis.

3.1.4. All-Embracing strategy

Despite earlier publications, Konrad Zerr (2003, p. 1) stresses not only the nature of the activities but rather delivers an overreaching definition.

"The term Guerilla Marketing indicates the selection of atypical and undogmatic marketing activities which, at a remote use of means, are supposed to deliver a preferably big effect. Guerilla Marketing has become a marketing mix overreaching basis strategy, a market political tenor of market penetration, which consciously searches for new, unconventional, until now ignored, possibly frowned upon possibilities outside the classical paths."

Zerr refers to a variety of activities, unbound to certain instruments or tools and stresses the dynamic nature of Guerilla Marketing as a basis strategy. Conceptually he assigns Guerilla Marketing a comprehensive position.

Zerr's definition has also been picked up by Nufer (2013). Nufer, who further complemented the definition by Zerr (2003) by referring to Levinson (2008) and Schulte (2007), confirmed by stating "Guerrilla marketing is an alternative, holistic marketing approach" (Nufer, 2013, p. 1). He also stated that "surprising content can turn a classical advertising vehicle or medium into a guerilla marketing activity" and that Guerilla Marketing "does not make traditional marketing obsolete, but has a supportive point-to-point effect and helps the marketing mix, mainly the communications mix, to take on an innovative new face" while aiming to gain maximum attention at relatively low costs (Nufer, 2013, p. 2).

Hutter and Hoffmann (2011a, p. 124) defined that *"Guerilla Marketing is an umbrella term for unconventional advertisement campaigns which aim at drawing the attention of a large number of recipients to the advertising message at comparatively little costs by evoking a surprise effect and a diffusion effect. In this way, guerilla marketing campaigns are highly efficient in terms of the ration of cost and benefit"*.

This definition bears various similarities to most recent contributions whereas Hutter and Hoffmann (2013) ignore the developments within the concept and do not stress the overreaching strategical view (even though they consider their importance and therefore classify them to be positioned in an all-embracing category). Interestingly Nufer also addressed the surprising characteristic of Guerilla Marketing moving his point of view even closer (Nufer, 2013). Apart from a subsequent Diffusion and Low Cost Effect, surprise is the fundamental subordinated effect through which Hutter and Hoffmann (2011b) have defined and modelled the Guerilla Marketing effect (see chapter 5). The Guerilla Effect, described as the ratio of benefit to cost, is also content of further definitions of Hutter and Hoffmann (2011b). Both display surprise as the starting point for the further effects and propose that without surprise and the emotions aligned, a Guerilla Effect cannot be achieved. These subordinated effects can be found in their definitions and are further being used to align principles and instruments and make up mediator variables in their models which are being discussed in detail in chapter 5.

After the presentation of the most relevant selection of definitions the presentation of the actual main issue constantly named, the so-called "Attention Dilemma", deserves further attention for processing. Only then an exhaustive critical discussion is reasonable, which can be found at the end of this chapter.

3.2. Attention dilemma

As a result of the information overload recipients discover on a daily basis, they naturally filter the mass of information and instinctively evaluate an advertisement to indicate its value (Hutter and Hoffmann, 2013). This lack of efficiency is not caused by the missing quality of the advertisements but rather caused by a strong antipathy of the consumer experiencing the messages as "disrupting and distracting" (Langner, 2007, p. 14). When consumers do not perceive advertising as worthy, or when excessively exposed, they tend to avoid it (Wanner, 2011).

In classical consumer communication, increasing competition, homogeneous products and the saturated markets that companies compete in, call for unconventional alternatives to reach the consumer (Ay et al., 2010; Moor, 2003). The lack of creativity in classical marketing communication leads to refusal and disclaim (Patalas, 2006). The reason why consumers are incapable of consuming whatever companies throw at them is not the quality of the marketing effort but rather the attitude of the recipient (Langner, 2007). The changes of the various factors do not call for companies to invest even more budget but, in fact, the opposite. Over 60% of any marketing campaign is said to have an insufficient cost-benefit ratio with the results simply not justifying the invested money (Langner, 2007).

The previous chapters already presented most of the major influential factors for the development of alternative marketing methods. The reasons described above, all lead to a final and decisive consequence described by Hutter and Hoffmann (2013, p. 6): "The Attention Dilemma".

At the starting point of this depiction the authors identified homogeneous products. Those homogeneous products put great pressure on the competing organizations which again react through increasing pressure on consumers by raising the amount of advertising messages (Hutter and Hoffmann, 2013). For the recipient this pressure results in an information overload. The overload is based on the fact that

humans are only capable of processing a limited amount of information over a certain period of time (Jacoby, 1977). Once capacities are met the perception of individuals is limited and decision-making processes are becoming imprecise (Jacoby, 1977). The reaction to experiencing an information overload is called "reactance" describing the antipathy towards the constraints put on the individuals (Hutter and Hoffmann, 2013, p. 6). Such constraint causes a negative perception of associated brands or products (Hutter and Hoffmann, 2013). Moreover a homogeneous communication of homogeneous products does not lead to success but rather to a so-called "wear-out-effect", stating that recurring of similar products, methods or campaigns causes boredom which again causing negative perceptions (Hutter and Hoffmann, 2013).

Finally, at the bottom of figure 3, the decreasing efficiency of advertising again seduces organizations to intensify advertising creating even more pressure (Hutter and Hoffmann, 2013). Hence, escaping the cycle can only occur when an organization decides to either decrease advertising efforts or implements alternative communication methods (Hutter and Hoffmann, 2013). Practitioners and scientists consider the Guerilla Marketing concept as an appropriate method to countervail the decreasing efficiency of classical modes of communicating through "attention recuperation, without evoking reactance" (Krieger, 2012; Hutter and Hoffmann, 2013, p. 7). From a consumer point of view the depiction of the Attention Dilemma described above delivers a straightforward explanation of prior and current developments in marketing. Yet, it also delivers a basic approach of how to leave this cycle of decreasing efficiency and is supposed to justify the move of marketers towards the so-called "alternative communication methods" (Hutter and Hoffmann, 2013, p. 6).

Having identified the fundamental problem statement and current approaches of defining Guerilla Marketing, the following chapter proceeds with the question of how alternative communication theoretically intends to overcome the attention dilemma, e.g. how to break the downward sloping efficiency curve of the classical approaches.

3.3. Résumé

As depicted above most current definitions refer to the possibility of overcoming the Attention Dilemma through the use of unconventional, alternative or non-traditional measures. Emotional activation (e.g. surprise) is named as the key to reaching strong effects and furthermore potentially efficiency through diffusion. Diffusion can be identified as the most relevant determiner for the efficiency in regard to costs, while only Patalas (2006) addresses the relevance of the two-step process communicated content moves along (e.g. first and second publicity). Apart from addressing how content is being forwarded, it is also vital to the relevance of how far content is supposed to travel and the consequences of the journey. "The further the better" can be accounted for most actions, yet the Guerilla Marketing instruments have different focus points and the further a message has travelled and the more people have noticed it, the less there are left to surprise anymore. Therefore the possibility of repetition is strongly influenced. Such distinction delivers also great support for a further separation between the target of Guerilla Marketing and its fit to either large or only SME's regarding the company's goals (Patalas, 2006).

Scholars widely agree that Guerilla Marketing can be recovered in the classical marketing mix with strong tendencies towards promotional tools (e.g. advertising) yet besides a few, they do not limit its radius and postulate the flexible nature in being adaptable to all four tools of the marketing mix (Hutter and Hoffmann, 2013; Krieger, 2012). A further point showing the developments over time can be found in whether or not Guerilla Marketing can and should be used in combination, isolation or as replacement for classical advertising (Levinson, 1984). Nowadays scholars propose the combination with classical measures instead of isolated use (Hutter and Hoffmann, 2013; Patalas, 2006; Nufer, 2013; Zerr, 2003).

Finally, so far little effort has been made to describe those Guerilla Marketing specifics which are supposed to be accounted as the origin (e.g. attention, surprise). In the end surprise as an emotion that desires cognitive processing it cannot simply be accounted as the trigger without further dedication (Hutter and Hoffmann, 2013). Those variables determining what surprise really constitutes are widely ignored. While stimuli (e.g. non-conservative marketing executions) are exemplary being addressed as origin and cause for surprise in further approaches, even the newest definitions seem to ignore the relevance. A gap between descriptive psychology and Guerilla Marketing sciences can be identified alongside the ignorance of behavioral inclusion.

3.4. Critique

Considering general marketing conceptions with the three steps of defining marketing goals, marketing strategies and marketing mix it is comprehensible that the procedure also applies to Guerilla Marketing concepts (Meffert et al., 2011). Furthermore scholars agree and prove that Guerilla Marketing in the marketing mix takes place primarily in the field of promotions, e.g. advertising, that it has become, or at least should be considered, an overreaching strategy and it potentially applies to any sized organization (Schulte and Pradel, 2006). Any further approach of narrowing down and categorizing does not grant exclusivity and primarily tries to show where the concept mostly, but not exclusively, fits. Exemplary defining Guerilla Marketing as only being applicable below-the-line or out-of-home is not sufficient and can easily be refuted as "invisibility" is not a circumventing characteristic (Hutter and Hoffmann, 2013; Krieger, 2012). Due to the range of possibilities any given advertising vehicle or mass-media communication tool could be integrated in a Guerilla Marketing campaign (Bruhn, 2007; Nufer, 2013).

Neither the tool of the marketing mix or instrument used, the conceptual operative planning, the predetermined surprise or activation, nor the desired effect, can be considered

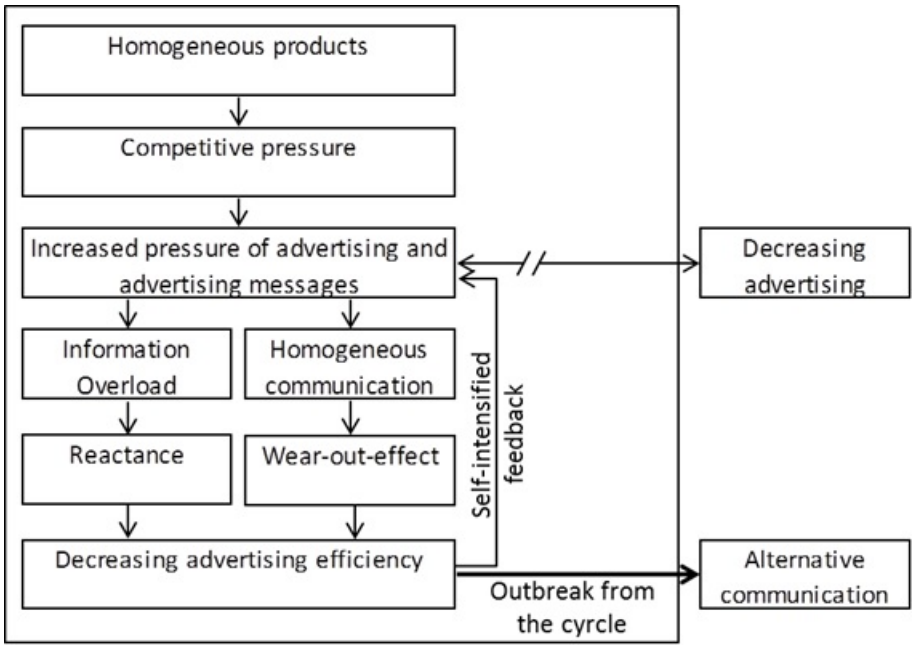

Figure 3: The Attention Dilemma; Source: following Hutter and Hoffmann (2013, p. 6)

an overreaching stand-alone criteria or exclusive characteristic (Meffert et al., 2011). Only in combination they may be. When targeting a common ground in Guerilla Marketing one has to critically question if there really is existence of delimitating characteristics and if this thin yet dynamic line will ever be identified. In the end, having filtered these proposed characteristics, one is left with a philosophical projection that is relying on the unconservative nature in the sense of altering classical advertising only in a way that assigns space to creative executions.

Ultimately the author proposes the following definition:

Guerilla Marketing is a comprehensive basic strategy executable throughout the marketing mix as a supplement, modification to and/or creative escape from any licensed and/or purchased mass-media advertising space. It is the art of executing strategical and economical war on competitors and framed spaces by generating cost-efficiency through achieving a variance in the audience's perception and expectation (e.g. surprise) to overcome the attention dilemma, with the potential to either trigger diffusion to a second extensive, yet relatively intangible and undefined, audience or to forcefully embed a relatively strong message in the first audience.

4. Guerilla Marketing practice

"A selective distinction and unambiguous categorization is not possible" and all Guerilla Marketing instruments "complement each other and function synergistically to produce the actual impact of Guerilla Marketing" (Nufer, 2013, p. 5). Therefore the upcoming section will present a selection of instruments that are widely acknowledged to be gathered under the umbrella of Guerilla Marketing. Besides classifying the instruments through concentrating them by their primary outcome or effect (e.g. the basis for the later presented mediators of the LSD Model I and II), the instruments described adjacently further correspond to certain principles (Hutter and Hoffmann, 2011b).

Nufer (2013) contributed a classification into 4 categories, whereas three are distinct and one, the "Low Budget Guerilla Marketing", is of a comprehensive nature. The other categories are labeled "Infection Guerilla Marketing" corresponding to those instruments, mostly new media, primarily aiming at a so-called Diffusion Effect (e.g. Viral Marketing), "Surprise Guerilla Marketing" corresponding to those with the emphasis on a strong Surprise Effect through installations in public places and lastly "Ambush Marketing" as a separate category (Hutter and Hoffmann, 2013; Nufer, 2013). The category of "Low Budget Marketing" is "a specific case that can be subsumed to all other categories" (Nufer, 2013, p. 3).

While Nufer (2013) and Hutter and Hoffmann (2011a) deliver fairly similar approaches those titled by Hutter and Hoffmann are used as depicted in figure 4. They close the gap between categorization, aligning instruments and fitting them under principles and effects that are later serving as a basis for modelling Guerilla Marketing. As depicted in figure 4, the principles presented by Hutter and Hoffmann (2011a) are mundanely categorizing the instruments of Guerilla Marketing into the principles of "Freeriding", "Living Environment" and "Recommendation". Freeriding corresponds to the Low Cost Effect, the Living Environment to the Surprise Effect and Recommendation, as a principle, aligns with the Diffusion Effect. As a fourth principle the "Flanking" Guerilla PR is presented, whereas its special frame will be discussed in detail later on (Hutter and Hoffmann, 2011a). While Hutter and Hoffmann list Ambush Marketing as a Low Cost Effect category, Nufer assigns a separated role to the concept

as it does not depend on Surprise or Diffusion Effects but can rather be executed without those features. Nevertheless both approaches make use of a similar foundation. Delineation only occurs while assigning the freeriding principle to the Low Cost Effect since Nufer states that "Low Budget Guerilla Marketing" as a category needs to be assigned to all other categories (Hutter and Hoffmann, 2011a, 2013; Nufer, 2013).

4.1. Guerilla Marketing instruments

In the following section the most relevant instruments will be presented, briefly discussed and the prioritized intentions will be presented. Both the desired primary effect of the Guerilla LSD Model I and II (see chapter 5) as a basis for the following chapters and the classification through assigning the relevant principle will support the subsumption. Simultaneously, issues and deficits of such classifications will be presented aside the examples. Comparing the instruments, if possible, to classical methods, will support the understanding (Hutter and Hoffmann, 2013; Nufer, 2013).

4.1.1. Ambush marketing

Ambush Marketing first arose before and at the Los Angeles Olympic Games of 1984. As the total number of official sponsors in the 1976 Montreal Summer Olympics peaked at a total of 628, regulatory changes introduced by the International Olympic Committee in the early 1980s restricted the number of sponsorships and partners of the Olympic Games from 1984 on (Shani and Sandler, 1998). A smaller circle of exclusive partners categorized by products was, on the one hand, supposed to deliver a greater awareness and stronger impacts for the sponsors and, on the other hand, drive the costs of these exclusive sponsorships (Shani and Sandler, 1998). Hence these changes opened the door for Ambush Marketing as a strategy for organizations not willing to pay these great amounts but still seeking the positive effects of being associated with major sporting events (Burton and Chadwick, 2009). As Burton and Chadwick (2009, p. 2) state, "sponsorship ambushing has historically taken place at major sporting events, and often has been seen as being directly related to competition between major brands (...)." Without officially being set in relationship to an event, these organizations make use of television commercials, social media, on-location promotions and out-of-home media strategically trying to be linked to the event and causing confusion with the recipients on whether or not an organization really is sponsoring an event or not. Meenaghan contributed a classification of five Ambush Marketing strategies: Sponsoring media coverage of an event, sponsoring a sub-category within an event, making a sponsorship-related contribution to a players' pool, engaging in advertising that coincides with a sponsored event, or the development of other imaginative ambush strategies (Meenaghan, 1994).

Studies have proven that Ambush Marketing and sponsorship both as concepts cause confusion with the recipients. The sponsorship concept is quite familiar whereas the Ambush Marketing concept isn't (Hutter et al., 2011). During the FIFA World Cup 2010 Hutter et al. (2011) conducted a study on how recipients notice and classify either sponsors or ambushers. Interestingly only 1% of the recipients were able to clearly identify the four tested ambushers whereas 74.9% of the recipients classified at least three out of four ambushers as official sponsors (Hutter et al., 2011). Contrarily only 64.4% of the recipients were able to reach the same score with the official sponsors while the classifications of the recipients prevailed that 63.3% of the recipients showed great confusion on who is an official sponsor and who is not (Hutter et al., 2011).

Ambush Marketing is one of the instruments with a high potential for experiencing negative consequences of various kinds. Apart from the danger of legal consequences, organizations can draw great profit out of ambush efforts as the costs (even including a legal aftermath) are relatively still much lower than those of an official sponsorship (Burton and Chadwick, 2009). Official right holder's investments are devalued and as control of the event organizers are lacking towards ambushing organizations, the risk of negative image transfers towards the events are increased (Burton and Chadwick, 2009). Yet literature, just like media and obviously the sponsors themselves link Ambush Marketing with a negative image and raise moral concerns, as parasitically enriching organizations not only harm the sponsors, but also cause damage to the events themselves (Burton and Chadwick, 2009).

Just as with other instruments, the evolution of Ambush Marketing has not stopped and changes in the execution and also the acceptance could quickly be discovered. Already in 1994 Meenaghan found that many of the actions and campaigns previously identified as Ambush Marketing were become routine practices.

Obviously Ambush Marketing uses the principle of Freeriding through profiting from publicity-effective events more than any other instrument. It carries its stand-alone image amongst others and could potentially be executed with the goal to reach a great mass of audience or simply a homogeneous group attending an event (Nufer and Bender, 2008). It is the guerilla version of an official sponsorship and primarily aims at achieving cost-efficiency (Hutter and Hoffmann, 2013). Furthermore it is not bound to any restrictions during implementation and could be anything from a logo on a shirt worn under a jersey, a big screen positioned on private housing right next to a stadium or just strategical advertising bookings with indirect references in the content to an event. Recalling a bilateral categorization from chapter 2 it becomes obvious how Guerilla Marketing campaigns either live off the concept or the instrument. Both are applicable and especially Ambush Marketing identifies itself rather through a war on licensing and competitors rather independently from vehicles or tools used, but rather through creative strategies. In Online-Appendix 1 a current example of the clinch between football sponsors and ambushers is presented whereas the scandal of Mario Götze wearing his personal sponsor's shirt received even more notice when the sponsor denied any responsibility for the action (see Online-Appendix 1).

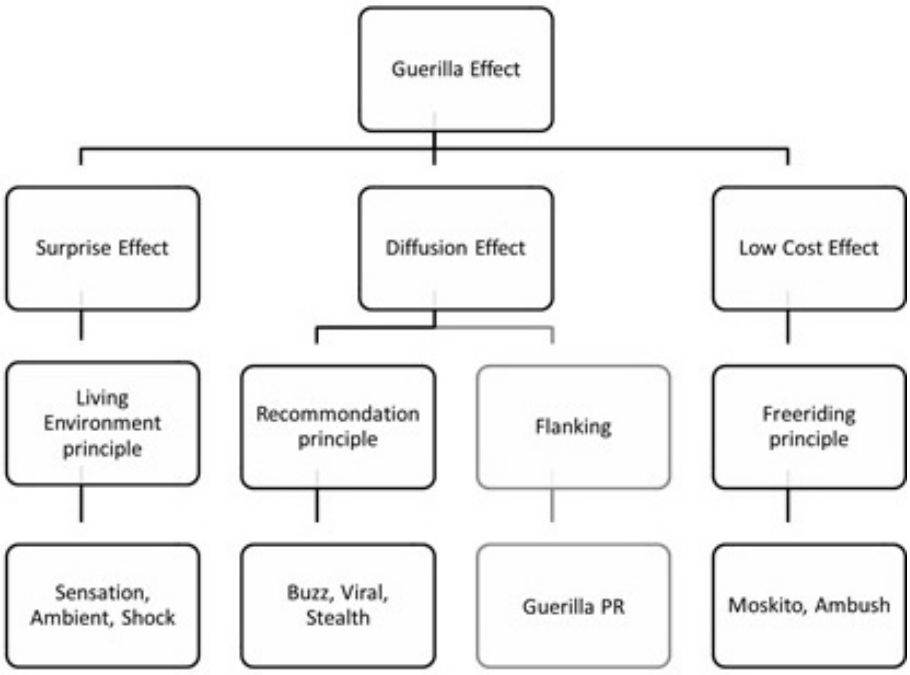

Figure 4: Alignment of Guerilla Marketing instruments, principles and effects; Source: following Hutter and Hoffmann (2011a).

4.1.2. Mosquito marketing

Just like Ambush Marketing, Mosquito Marketing has the ultimate goal of receiving attention through apportion of other's costs (Hutter and Hoffmann, 2013; Ries and Trout 1986). While Ambush Marketing is frequently performed by huge organizations, Mosquito Marketing is rather performed by SME's which are trying to fill a big competitor's niches (Patalas, 2006). The effect of the marketing efforts is also considered a Low Cost Effect through making use of the freeriding principle whereas in this case competition could allow for or even profit from it. While Ambush Marketing seeks mixed perceptions and confusion amongst the consumers, Mosquito Marketing is rather about positioning and aiming a niche target group often by offering complimentary goods or services and promoting them (Ries and Trout, 1986). Jack Trout, a pioneer in market segmentation, primarily pushed the concept towards an understanding of Niche Marketing understanding the concept as an Assault Strategy as discussed in chapter 3 (Ries and Trout, 1986).

4.1.3. Shock marketing

As the first instrument mainly working through the stimulation of a strong Surprise Effect and making use of the living environment principle, Shock Marketing is being introduced (Hutter and Hoffmann, 2013). This method tries to use shock through delivering campaigns violating social norms, values and laws, at an extreme emotional state and is therefore said to stimulate a maximum effect (Williams, 2009). "Emotional Marketing is able to imprint itself over a longer period of time and easier into the memory of the recipient." (Nufer and Bender, 2008, p. 19). Nevertheless, shock is a very subjective state of mind and therefore various content can and will be perceived differently by whoever consumes it, whereas many campaigns simply show a non-disarmed reality that the recipient is not capable of handling (Williams, 2009). Whereas optimists would suggest leaning on the slogan that "any publicity is good publicity" only very few organizations dare to use the instrument of Shock Advertising fearing a negative image transfer (Williams, 2009).

Conversely recipients tend to ignore and refuse shocking advertisements as a self-protecting mechanism (Williams, 2009). Any content that contradicts morals, values, norms or religious beliefs like for example extreme nudity, racism, brutality, anti-Christianity or even poverty embedded in the desired surrounding will most likely draw great attention through a high level of activation, yet the reactions of the individual are hardly to be foreshadowed and can quickly result in an antipathy towards an associated brand, product or person (Williams, 2009) Only if that is the desired reaction, an initiator can run a Shock Advertising campaign at barely any risk. Legal regulations for example in some countries enforce the placement of shocking pictures and messages on merit goods like tobacco products or liquor. In this case shocking the recipient in a way that supposedly discourages the use and enhances antipathy towards the product is desired (Urwin and Venter, 2014). Another example is the brand "French Connection" that was found in London in the early 1970s. Their great marketing success from the early 2000s till today was also caused by the company's implementation of the initials "FCUK" meaning "French Connection United Kingdom". The equivocal acronym together with an adverting slogan drew great attention and triggered a provoking yet

appealing reaction reaching the right target group (Javed and Zeb, 2011).

4.1.4. Ambient marketing

Ambient Marketing is a concept which, as the name suggests, mainly deals with advertising in a specific living environment or ambience and which surprises through leaving frames of mass-media advertising (Bruhn 2013; Hutter and Hoffmann, 2013). The British agency "Concord Advertising" started promoting it in 1996 and is mainly responsible for its breakthrough as a reaction of their client's requests for something new and different (Chatterjee, 2011). The terms "Ambient Marketing" or "Ambient Advertising" are commonly used synonyms whereas for the sake of distinctiveness and despite the fact that its founder Concord Advertising titled the concept Ambient Media, the label "Ambient Marketing" will be used (Wehleit, 2005). As Ambient Marketing can hardly be assigned to an advertising vehicle or frame, common definitions only rarely try to identify aligned media but rather focus on the interplay of the ambience and the target groups (Grauel, 2002). An example is Ralf Grauel's definition stating that Ambient Marketing as a "non-classical media communication with mobile target groups" whereas again still today no comprehensive definition has been recognized or acknowledged (Grauel, 2002, p. 2). It can further be classified as an out-of-home-media and can be seen as the Guerilla Marketing version of the classical outdoor-advertising (Hutter and Hoffmann, 2013). Borders shift over time and a clear classification towards a specific outdoor-advertising or setting a clear distinction between classical and non-classical advertising is almost impossible. Secondly "Ambient could be seen as a movable term and defined by advertising norms of the day" (Luxton and Drummond, 2000, p. 735). For example the distribution of free post cards with advertising messages in bars has become a classical mass-media medium within ten years whereas at the beginning it was a non-classical innovative form of Ambient Marketing. Just like any abnormal location or innovative idea is considered different, "they are perhaps no longer unconventional and have encroached on mainstream territory by virtue of their repetition" (Luxton and Drummond, 2000, p. 735). Welling (2005) contributed a broad classification of Ambient Media following the features of indoor vs. outdoor, unique campaign vs. enduring campaign and mobile vs. fix installations.

Whether or not the concept itself is an extension of the usage of advertising space or an innovative concept is broadly discussed, but in the end it is aiming at overwhelming the consumer with an advertisement in an unexpected situation trying to maximize attention through surprise (Chatterjee, 2011). The features characterizing an ambient campaign are therefore choosing the right, e.g. unexpected and unusual, location, the unconventional method of positioning or executing an advertisement and the time-dependency, e.g. the temporal factors (Luxton and Drummond, 2000). The multifaceted field of application opportunities again support a bilateral perspective as introduced in chapter 2 and 3. Most relevantly the ambience, e.g. the atmosphere or habitat, of the specific target group justify Ambient Media (Wehleit, 2005). This could be the supermarket, tram or simply a bar whilst the delimitation of a homogeneous target group towards others could in this case turn a simple poster into an Ambient medium (Wehleit, 2005). As mentioned before this interplay only works as long as Ambient Marketing still comprises the most relevant Surprise Effect, the deviations of schema-incongruence, as a key descriptor (Croft, 1998). The importance of the homogeneous target group therefore obviously determines the projectable characteristic of any definition (Wehleit, 2005). In summary one needs to notice, that "what might be considered ambient one day may not be the next, as audiences become familiar with any given location and the point of difference or 'unusuality' fades" (Luxton and Drummond, 2000). The overall aim remains the attempt to cut through the clutter of conventional advertisement by disrupting the consumer and pleading him or her for attention (Bainbridge and Curtis, 1998).

Due to the relatively low cost, the wide range of opportunities and the potential of diffusion to a great mass of recipients, Ambient Marketing is one of the fastest growing instruments (Turk et al., 2006).

4.1.5. Sensation marketing

The term Sensation Marketing is one that in practice is often used as a direct synonym for Guerilla Marketing as the campaigns and actions usually entail a high effort and degree of singularity which makes it almost impossible to replicate (Nufer and Bender, 2008). Embedding a campaign in a strategy that enhances a strong Diffusion Effect is therefore the factor determining the success, whereas embedding the campaign in the living environment and therefore making use of the Surprise Effect is the most relevant classification factor (Hutter and Hoffmann, 2013). Surprise sets the condition for diffusion. The actions are usually executed in the out-of-home area in highly frequented regions and locations promising to reach a great mass of recipients and the media for just a short period of time (Schulte, 2007). The delimitation to Ambient Marketing results from the dynamic character of Sensation Marketing whereas Ambient Marketing is characterized as static and therefore less restricted by time (Hutter and Hoffmann, 2013). Dynamics are mainly generated by persons being involved (Schulte, 2007). Sensation Marketing further confines through bearing greater and different risk factors of failure through the uniqueness of an event, the greater dependency on a strong Diffusion Effect and risk of missing the target groups. Often concerts, tourist attractions and sporting events are therefore used as a setting to be able to forecast a potentially homogeneous target group.

4.1.6. Buzz and stealth marketing

While Viral, Buzz and Stealth Marketing all work through the positive effects of WoM and recommendation principles and therefore can be classified as prioritizing Diffusion Effects, they are often mistakenly used as synonyms (Hutter and Hofffmann 2013).

Buzz Marketing can be described as *"a multi-dimensional communication process that involves sending persuasive messages out via buzz agents (senders) to buzz targets (receivers) in the form of WOM (offline and online) conversations and retrieving market research information from buzz agents on the particulars of these buzz marketing experiences"* (Ahuja et al., 2007).

As Buzz Marketing simultaneously works through quickly spreading content online and offline amongst communicating individuals, most scholars agree on delimitation to Viral and Stealth Marketing through the use of so-called "Buzz agents" as facilitators (Ahuja et al., 2007; Mohr, 2007; Thomas, 2004; Kaikati and Kaikati, 2004). Most importantly those agents act voluntarily as commissioners of a brand, organization or product whereas Stealth Marketing makes use of interlocutors purposefully spreading a concealed message (Kaikati and Kaikati, 2004). It is the credible propagation of something the agent is enthusiastic about versus the "economically motivated" Stealth approach (Hutter and Hoffmann, 2013, p. 84). Despite actual enthusiasm the agents (often mediated via Buzz agencies) can still get paid while taking the role of the consumer, opinion-leader, brand commissioner, sales supporter or market researcher (Ahuja et al., 2007). Both approaches raise moral concerns whereas critics fear a commercialization of the private life (Hutter and Hoffmann, 2013). Whilst Buzz agents can freely decide on whether or not they want to lay open their function, Stealth Marketing purposefully hides messages and the commercial roots of the communication and therefore calls many critics trying to prohibit Stealth Marketing (Walker, 2004). Most relevantly the general concept itself only works through those individuals that can either be motivated intrinsically or extrinsically to take the role of an agent (Walker, 2004). Those who like communicating, sharing information and strongly involve themselves in brands or products are intrinsically motivated whereas extrinsic motivation can be of an immaterial nature like prestige or a material nature like product samples or payments (Walker, 2004; Hutter and Mai, 2013).

For the sake of integrity it is crucial to understand the misleading use of terminology and category. Despite the synonymous use of Buzz and Stealth Marketing, various scholars use Stealth Marketing as the umbrella term for a great variety of other methods and types of Marketing (Kaikati and Kaikati, 2004). Kaikati and Kaikati (2004, p. 6) deliver a distinction between Buzz and Stealth Marketing saying that "instead of aggressively shouting to everybody at the same time, stealth marketing tends to whisper occasionally to a few individuals." Andrew Kaikati, a consultant serving Fortune 500 clients, and Jack Kaikati, Professor of Marketing at the Southern Illinois University (Kaikati and Kaikati, 2004, p. 6 ff.), also list the following techniques as subordinated types of Stealth Marketing: Viral Marketing, Brand Pushers, Celebrity Marketing, Bait-and-Tease Marketing, Computer Game Marketing and Pop/Rap Music Marketing opening the field of interpretation even further yet giving another example of the special role of Viral Marketing as a supportive tool.

4.1.7. Viral marketing

Picking up Viral Marketing and considering that it is the intention (as the name suggests) to purposefully spread content amongst individuals, one has to regard the bipolar use of the concept. First of all it is being used in isolation to cost-effectively reach a great mass of recipients but secondly it is often aligned with other advertisements like Ambient Marketing to facilitate a maximum Diffusion Effect towards a great mass of recipients. Nevertheless the diffusion through positive WoM and recommendation is the desired effect and therefore provides the basis for its classification (Hutter and Hoffmann, 2013).

The concept strongly depends on word-of-mouth transfers, one of the oldest and still most relevant marketing techniques (Röthlingshöfer, 2008). In this context WoM can be defined as "(...) any sort of communication between individuals about products, services, brands which takes place without a commercial interest (between the individuals). It is not necessarily bonded to oral or written comment. The participants or word-of-mouth can further transfer information via pictures, videos or through any communication method" (Röthlingshöfer, 2008, p. 27). The art of trying to regulate this interpersonal exchange of individuals strongly depends on so-called "Memes", which are ideas of any kind which are imprinted in an individual's mind who then acts as a host to transfer the content just like a virus (Langner, 2009, p. 15). Only those Memes that are enhanced through catalysts called "memetic triggers" will be forwarded to infect further individuals (Röthlingshöfer, 2008).

The elements determining the success of a Viral Marketing campaign are the content, the so-called "seeding" and the incentives for the infected individuals to spread the information (Langner, 2009, p. 38). Firstly the content, meaning the product promoted and rather also the message communicated needs to be designed in a way that it can become a topic amongst the target group. It needs to entertain and amuse the consumer, trigger emotional reactions, deliver a positive image that the recipient is keen on spreading but mostly be innovative and new so that the consumer believes he or she is in an advanced situation being able to spread new content (Langner, 2009). An assigned use (e.g. sales promotions) can also serve as reason to spread. Being able to create a strong Surprise Effect can be a facilitator again (Langner, 2009). Especially mass media like the internet offers a wide variety of opportunities (Leonardi, 2008). Social networks, blogs, chats, newsletters and web communities within those networks allow the initiator to speak to a homogeneous target group (Leonardi, 2008). The range of seeding can be determined by either aiming at a flow of information towards close friends and relatives of a recipient (mainly relying on a high value of the content) or at extensive seeding towards a rather anonymous and random group of recipients (Langner, 2009). The incentives to forward any content can be of a passive nature (lowering barriers to share information) through for example enabling an individual to share information by just clicking on a link on any homepage

or of a more active nature offering discounts, promotions or competitions for those who involve others (Langner, 2009).

4.1.8. Guerilla PR

Guerilla PR can neither be directly classified as an advertisement vehicle nor a strict Guerilla Marketing instrument, but rather a "flanking" tool that may supportively accompany other Guerilla Marketing instruments (Hutter and Hoffmann, 2011b, 2013). Strongly depending on Surprise Effects within editorial content that are worthwhile being covered, Guerilla PR, considered a subtype of PR, can enhance Diffusion and Low Cost Effects and therefore be associated and combined not only with the different instruments but also with any subordinated effect or instrument of the Guerilla Marketing (Hutter and Hoffmann, 2013). Hutter and Hoffmann (2011b, p. 7) further state that "Guerilla PR markedly improves the Guerilla Effect since media reports increase the number of addressees at hardly any financial efforts" and "is very persuasive, because consumers consider press reports more credible than advertisements. Purposeful initiations can be achieved through the execution of practically any campaign that is planned "in such a manner that they are appealing for the journals" (Hutter and Hoffmann, 2011b, p. 7).

Yet again controversy is great and confusion the normal course. Marketing and journalism literature either ignore the concept or promote it as the umbrella term for any instrument or campaign that can potentially be covered by Guerilla PR (Fischer, 2009, p. 380 ff.).

4.2. Résumé

With the presentation of the individual instruments it becomes visible where the assessment by Nufer (2013) from the beginning of chapter 4 came from. Due to the extreme fundamental discrepancies between the instruments the discussions on categorizing and defining Guerilla Marketing are now even more understandable. Each of the three approaches introduced by Nufer (2013), Gelbrich et al. (2008) and Hutter and Hoffmann (2013) just like the assignment of principles are comprehensible in a segregated consideration. Yet, all instruments clearly show distinct individual characteristics and therefore an isolated perspective is reasonable. Nevertheless, even though fitting the instruments in categories is possible to a certain extent, those categories again will ideologically drift away from each other opposing approaches to finding an all-embracing approach.

4.3. Critique

Critical observation only is legitimate when considering the principles and effects individually in objection to the usage as a basis for defining Guerilla Marketing as a whole. Most strikingly as shown in figure 4 Hutter and Hoffmann (2011a, 2013) try to use categories derived from desired effects of the instruments (Surprise Effect, Diffusion Effect and Low Cost Effect) whereas, without forestalling chapter 5, those are later set in subsequent order for modelling Guerilla Marketing. Such ordering contradicts an approach to use the same categories for clearly delimitating the instruments, yet acknowledges a strong interdependence between the effects and subsequently the instruments.

Such interdependence, especially in regard to Viral Marketing and Guerilla PR, seems quite comprehensible. Those instruments should be considered as facilitators that bear the primary intention to enhance the effect of other instruments. In fact Viral Marketing is closer to being a transmitter, e.g. a new channel or vehicle, rather than an instrument. It has emerged from modern technologies offering further opportunities just like radio advertising or television advertising have at the beginning stages of their presence yet does not set any limitations to the form of content that is supposed to become a virus and allows for unintentional spreading. Guerilla PR is similar in this respect and, in a simple thought, just like individuals diffusing a virus a broadcaster would cover and forward content they perceive as worthy and somewhat editorial. Furthermore the contributions of Nufer (2013) and his idea of considering Low Budget Marketing as an overreaching principle that accounts for any instrument, such category seems rather self-explanatory applying basic economic principles.

Aligned with the (re-)introduction of a "Guerilla Marketing principle" overreaching all potential marketing tools or instruments (see chapter: 2.5), the author therefore proposes a general alienation from the instruments which mainly exist as such due to the frequency of execution in practice. Exemplary, if any instrument can potentially become a Guerilla Marketing instrument, then assigning new titles for a Guerilla Marketing "version" of another instrument is little constructive. Ambush Marketing is the guerilla version of any other advertising or sponsoring of an event. Shock, Ambient or Sensation Marketing can be perceived as advancements in the evolution of marketing due to the need of reacting to an Attention Dilemma (see chapter 3). Buzz and Stealth Marketing are an extreme combination of WoM driven by customer loyalty and intrinsic and extrinsic motivation whereas Viral Marketing and Guerilla PR are primarily transmitters (e.g. channels) with supportive functions (Walker, 2004). Finally Guerilla Marketing can still be described through its instruments, yet it should never be limited by those instruments or whatever title they carry. The dynamic character is what keeps the principle alive while instruments come and go never forgetting how the title was linguistically derived (Nufer and Bender, 2008).

5. Modelling Guerilla Marketing

Despite the issues regarding the formulation of an up-to-date definition, even fewer approaches have been made to present an overall model for Guerilla Marketing. While few approaches aim at modelling individual Guerilla Marketing instruments (e.g. "R.I.P. Conceptual Model for Ambient Advertising" by Snipes, 2012; "Alternative Marketing Model on Social Media Platforms" by Castronovo and Huang, 2012), most relevant all-embracing contributions over the last years have been published by Hutter and Hoffmann

(2011b, 2013). Furthermore scholars like Luxton and Drummond (2000, p. 737) made use of classical marketing models (e.g. "Attitude-toward-ad-model" by Edell and Burke, 1987) trying to integrate Guerilla Marketing measures (e.g. "Ambient Model" adapted from Sheth-Newman-Gross-model of consumption values 1991). Basic models like the "AIDA" or "S-(O)-R-Model" (Gawlik et al., 2006) serve as a ground for subsequently organized approaches.

While the models presented were derived from current approaches of classifying and defining Guerilla Marketing, they will serve as a basis for the upcoming chapter.

5.1. LSD Model I

Aligned with their definition, Hutter and Hoffmann first presented an approach to model Guerilla Marketing which is titled "LSD Model" referring to the three mediators positively influencing the outcome: "Low Cost-, Surprise- and Diffusion Effect" (Hutter and Hoffmann, 2011b, 2013). Definition and model both depict that Guerilla Marketing lives on achieving a so-called "Guerilla Effect" describing, the favorable ratio of cost and benefits, with the goal of a profitable relation of advertising costs and effects (Hutter and Hoffmann, 2011b). This Guerilla Effect (see figure 5), modelled as the dependent variable, is being influenced through the independent variable, an "unconventional advertising campaign" (Hutter and Hoffmann, 2011b, p. 4).

Apart from the relationship between the independent variable, the three mediators show causal relationships between the dependent and independent variable. The total effect of the model is therefore explained only via indirect (mediator) effects, whereas no direct effect is depicted (Urban and Mayerl, 2007). The three subordinated effects, the mediators, each either influence the dependent variable directly or the subsequent mediator if applicable. Multicollinearity exists. Yet as the mediators explain why and how a causal relationship exists and since the model from 2011 has undergone various changes regarding the mediating variables, focus should be put on the three subordinated effects and the overall understanding of the model (Hutter and Hoffmann, 2013; Little et al., 2007, p. 207 ff.).

The designated benefits of the Guerilla concept mentioned earlier frequently describe the "increase in consumer's attention" as one of the few points most definitions agreed on (Hutter and Hoffmann, 2011b; Ries and Trout, 1986; Levinson, 1984). "Attention", the variable of the ratio that needs an increase to optimize the measure, should be provoked through using surprising elements in Guerilla campaigns (Hutter and Hoffmann, 2011b). In succession the "Surprise Effect" is the first of the three subordinated effects creating the Guerilla Effect (Hutter and Hoffmann, 2011b, p. 4). To further stimulate an efficient ratio the cost-variable can be optimized through being able to reach a great mass of recipients. A transfer of a Guerilla campaign to a larger amount of recipients can therefore lead to the "Diffusion Effect" which stands in line as the second subordinated effect of the model (Hutter and Hoffmann, 2011b, p. 4). Interestingly and not depicted in the model this effect can be driven either by interpersonal diffusion or medial diffusion as the media could potentially pick up campaigns that drew their attention (Hutter and Hoffmann, 2011b). Lastly the so-called "Low Cost Effect" in a third step can be reached through executing campaigns that require only small budgets or that are able to relatively increase contact numbers reducing the cost per person.

Below those three subordinated effects of the LSD Model I will be explained in further detail followed by a transfer into the second model.

5.1.1. Surprise effect

"Surprise is a result of the divergence of perceptions and expectations" (Hutter and Hoffmann, 2011b, p. 4). Expectations are the result of patterns that an individual develops while a discrepancy between a stimulus (e.g. a Guerilla campaign) and a schema (e.g. fixed spots for advertisement that an individual expects to see) can cause surprise (Mandler, 1982; Alden et al., 2000). Surprise, as an emotional reaction to a certain level of disconfirmation, is the emotional reaction which subsequently leads to a change in cognitive activation, e.g. greater attention towards a stimulus (Meyer et al., 1991; Lindsley, 1951). Surprise can therefore be achieved by any stimulus which fulfills these requirements and which has the strength to emotionally influence an individual through exciting or shocking the recipient or simply, as mentioned above, by positioning advertisements in atypical locations (Meyer et al., 1991). Path 1 depicts the causal relationship of the advertising measure and surprise. Path 2 subsequently displays how the surprised individual shows increased and overproportioned interest in the stimulus achieving an indirect Guerilla Effect via only one mediator variable (Niepel et al., 1994). The instruments primarily focusing on a maximized Surprise Effect are Ambient Marketing and Sensation Marketing (Hutter and Hoffmann, 2011b). With Ambient Marketing surprise is generated through the use of unusual settings for implementing unexpected and unconventional campaigns and through being the first to do so. Usually those are positioned and bound to remain static for a certain period of time. Sensation Marketing on the other side tries to surprise through a more dynamic character while performing actions rather than permanent installations. Exemplary "flash mobs are a subtype of sensation marketing which is more and more often used" as public initiation (Hutter and Hoffmann, 2011b, p. 5-6). Hence such initiation shows how a strong Surprise Effect can generate the attention of individuals who would elsewise try their hardest to avoid advertisements in any way.

5.1.2. Diffusion effect

The amount of recipients reached with any campaign is classically the most determined matter of expense computed through the cost-per-thousand parameter in radio, TV or print media. Whilst trying to reach the maximum amount of recipients at given costs, Guerilla Marketing campaigns aim at profiting from multipliers, such as individuals or the

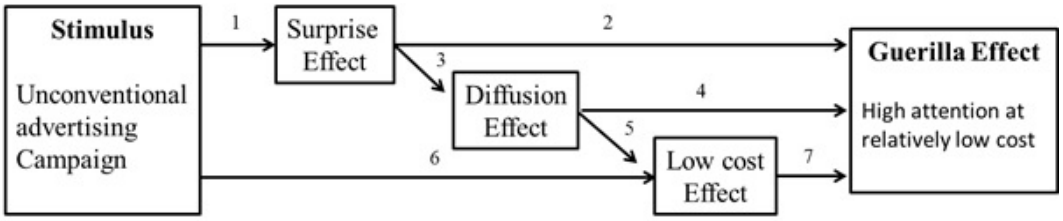

Figure 5: The LSD Model I; Source: following Hutter and Hoffmann (2011b, p. 4)

media sharing information of a campaign, that do not increase costs in a ratio the cost-per-thousand would suggest (Hutter and Hoffmann, 2011b). Hence it is vital to take as much control over these gratuitous multipliers as possible (Solomon et al., 2011). For example, especially through social networks, the initiators can try to evoke surprise in a way that constructively stimulates the voluntarily forwarding of content by recipients through emotionally involving them in a way that makes them obliged to share information (Menon and Uddin, 2010). This modern form of word-of-mouth communication, just as the classical, also bears the advantage of a high credibility as individuals perceive friends or family members as a much more reliable source entailing a credibility that no other classical advertisements could ever achieve (Hutter and Hoffmann, 2011b; Arora, 2007). Path 4 in figure 5 depicts how the surprise, as result, is being carried on through the Diffusion Effect creating the Guerilla Effect directly. Yet, a Low Cost Effect can also be the third mediator involved (path 5) if such diffusion occurs at relatively low costs.

Instruments that can primarily be assigned to the Diffusion Effect are Viral Marketing, Buzz Marketing and Guerilla PR. Viral Marketing mostly focuses on sparking a flow of communication that spreads like a virus into the desired target groups (Southgate 2010). Again the spreading of a content can be provoked either through individuals, e.g. members of the target group who hopefully forward the content to other members of the homogeneous group, or the media. Just like the flash mob instrument described above, Guerilla PR focuses on modern information technologies and communication technologies and most frequently uses short video clips that are supposed to diffuse through being shared over the internet (Hutter and Hofmann 2013). This way Viral Marketing is often described as a cost-efficient pendant to conventional television-spots but also serves as a facilitator to other instruments (Hutter and Hofmann 2013).

Buzz Marketing also follows the overall goal to stimulate online and offline word-of-mouth communication. Individuals, often fans of the product, are equipped with free samples and detailed information and should then act as "market researchers, advertisers, distributors and influencers in the decision-making process of other consumers" (Hutter and Hoffmann, 2011b, p. 7). Obviously these individuals need to be incorporated in a homogeneous target group with similar characteristics to be able to be persuasive (Hughes, 2005).

Guerilla PR, as a contraire to Viral Marketing and Buzz Marketing but as a facilitator to any other Guerilla Marketing instrument, focuses on making use of editorial sections of mass-media instead of accounting on individuals communicating (Hutter and Hoffmann, 2011b).

5.1.3. Low Cost effect

The third subordinated effect can either be triggered through cost-efficient unconventional advertising campaigns directly (path 6) or through the prior subordinated Diffusion Effect depicted by path 5 (Hutter and Hoffmann, 2011b). The foremost intention in this stage is to obtain the maximum output of a campaign with only investing little or no budget at all either through the diffusion via individuals or mass media or the Guerilla campaign itself simply needing barely any budget to execute (Hutter and Hoffmann, 2011b).

At this point the practitioner can potentially decide on whether investing a larger amount of money into creating a Surprise Effect gaining tremendous attention amongst individuals and the media and hoping on the message to diffuse or investing a smaller amount, or even no money at all, to initiate an effect of a smaller dimension. This could come free (one could speak of a "No Cost Effect") but will most likely draw less attention and therefore show a weaker Diffusion Effect (Hutter and Hoffmann, 2011b). Yet, a campaign of a smaller scale and without barely any budget invested could still be carried on through media to great dimension if it ignites. Nevertheless the intended results obviously depend on the initiator and his or her intention. A big organization would most likely aim at maximizing the Diffusion Effect to account on achieving a low cost-per-thousand parameter through reaching a great mass of recipients whereas a SME with small marketing budgets would potentially rather aim at minimizing costs. Both perspectives account on the Low Cost Effect.

Advertising space could not be more costly than during large events and in relation to the costs to purchase these licenses, no other temporary happening could deliver the attention of great events like the annual NFL Superbowl or the FIFA World Cup which is held every four years (Seguin and O'Reilly, 2008). Therefore sponsors are willing to pay vast amounts of money to be able to advertise and sell their products during the events hoping for further positive effects like image transfers and positive associations (Hutter and Hofmann 2013). Ambush Marketing primarily picks up exactly those costly events with the intention to profit from them on any level without investing any budgets for licenses or legal

permissions (Seguin and O'Reilly, 2008). Hence, it can be seen as the Low Cost Effect version of a classical sponsorship. Apart from using Ambush Marketing as a parasitic method to transfer a positive image companies also tend to attack direct competition with their ambush campaign intentionally weaken official sponsors (Hutter and Hofmann 2013).

5.2. LSD Model II

Two years after the first presentation of the LSD Model I of Hutter and Hoffmann a second version followed in 2013. Firstly mediator variables, the same as displayed in the LSD Model I, are now categorized by either enhancing the utility of a Guerilla Marketing campaign or having an effect on the costs whereas obviously only the Low Cost Effect is assigned to determining the costs (Hutter and Hofmann 2013). Secondly the authors rearranged the mediator variables, eliminated the independent variable ("Unconventional advertising campaign"), e.g. replacing it indirectly by "Use of a Guerilla measure", and limited influencing paths to a total number of five instead of the original seven (Hutter and Hofmann 2013). The Surprise Effect is now said to stand in direct positive correlation with the Diffusion and Low Cost Effect yet it can no longer directly influence the outcome variable and is neither influenced by the erased independent variable. The Diffusion Effect still bears the same direct effect on the dependent variable and an indirect effect mediated over the Low Cost Effect (see figure 6).

Moreover the individual subordinated effects are now bearing further explanations. The Diffusion Effect now is considering diffusion via two levels (stimulating WoM and stimulating media) and the Low Cost Effect entails relatively low costs in reference to low production costs and low contact costs per person (Hutter and Hofmann 2013). Most strikingly and presupposed through the former independent variable of "Unconventional advertising campaign" the Surprise Effect gained in importance. Schema incongruity is referring directly to the emotional reaction to a stimulus which is then causing a disruption and the drawing of attention (Hutter and Hofmann 2013). Upstream to the surprise the use of a Guerilla measure is listed sporadically.

5.3. Résumé

Various models try to display the functioning of individual instruments while only few depict Guerilla Marketing as a whole. The two models presented strongly rely on the three effects of Surprise, Diffusion and Low Cost whereas they serve different mediating purposes. Yet, the different paths and impacts cause big variances between both models while key paths have been removed or complemented. An upcoming critical review shall serve further explanation.

5.4. Critique

First and foremost the LSD Model I can be acknowledged as a first approach but shows various weaknesses. Firstly the most relevant issue is the reference to "unconventional advertising campaign" as independent variable. Most scholars (including the editors of the model) agree that Guerilla Marketing is not limited to advertising itself but can be used overreaching the entire marketing mix (Schulte and Pradel, 2006). Secondly the term "unconventional" is not finally described. As discussed in chapter 2 simply circumventing between media-advertising and the residuals cannot be sufficient and as the authors further only state that "unconventional" refers to being surprising, positioned at random locations and methods that do not deem to cause reactance, too much space for interpretation is left (Bruhn, 2007; Egan, 2007). Apart from this, the causal relationship between unconventional advertising and surprise is left uncommented. Such fragile emotional state deserves more attention. Moreover the Diffusion Effect does not refer to what Patalas (2006, p. 45) called the "first publicity" and "secondy publicity" and therefore lacks clarity. Furthermore objection occurs since the authors define Guerilla Marketing as strongly or only depending on a Surprise Effect whereas the model shows the opportunity to generate a Guerilla Effect without integration of the mediator variable of Surprise seen in path 6 and 7 (Hutter and Hoffmann, 2011b). Lastly a Surprise Effect has no direct path to the Low Cost Effect limiting the model to only allowing a Low Cost Effects to occur via Diffusion Effects and not directly through surprise.

The LSD Model II erased some of the weaknesses presented beforehand yet some new issues arose. Firstly and in line with the current definition, the origin of the second model now is the Surprise Effect yet the actual independent variable as stimulus is missing. The Surprise Effect now sporadically represents the independent variable. The "Use of a Guerilla Measure" is listed as a heading yet has no path showing a relationship to the Surprise Effect, while the reference to "unconventional" and advertising has been erased from the model. Hence moderators of the model showing the "changing of a relationship as a function of some moderating influence" are simply missing (Little et al., 2007, p. 216). Identifying schema incongruity perfectly describes the process along drawing attention, distraction and the development of surprise yet it is not fully describing the entire mental process (Meyer et al., 1997). Following given theory verification of given discrepancies, the assessment and evaluation (e.g. cognition) are ignored so far (Meyer et al., 1997). Now the origin of schema discrepancies needs revision due to their subjectivity or at least should be represented by a moderator from erased independent variable. Referring to the discrepancies also does not constrain the model or bear exclusivity towards e.g. media-advertising (Nufer, 2013). If the "Use of a Guerilla Measure" as the independent variable was only mediated via the Surprise Effect, the model would list a direct effect between an independent variable and mediator variable, yet such relationship is not clarified (Hutter and Hoffmann, 2013). The way it is depicted would then also suggest another independent variable to be present (e.g. "Costs of a Guerilla Measure"). Problems may also arise through listing "low production costs" as a characteristic of the Low Cost Effect. As the costs are relative and while Guerilla Marketing

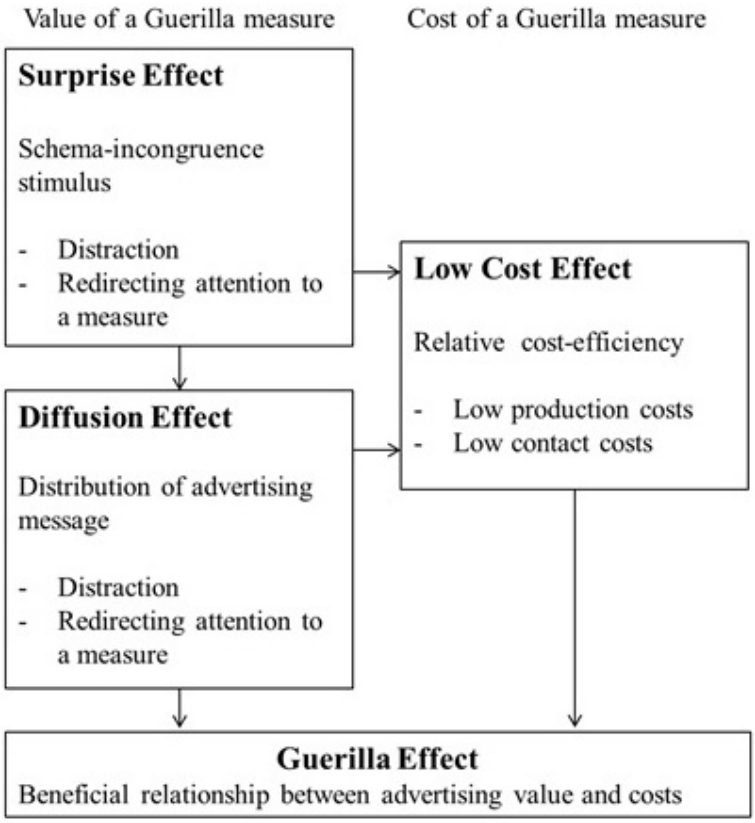

Figure 6: The LSD Model II; Source: following on Hutter and Hoffmann (2013, p. 36).

can either be executed through maximizing output at fixed input or minimizing input at fixed output, such projection can be fatal. Furthermore the controllability of the diffusion needs revision.

Guerilla Effect Model III: While some of the issues of the LSD Model I have been tackled (e.g. a two-step Diffusion Effect, the absence of the opportunity to generate a Guerilla Effect without a Surprise Effect) in the second model presented, some still need to be reviewed (Schulte and Pradel, 2006). The author therefore proposes a third model described below (see also figure 7).

Like in the LSD Model I, the presented Guerilla Effect Model III, shows a clear independent variable ("Guerilla Marketing") which is the execution of any possible Guerilla Marketing instrument or principle and the "Guerilla Effect" as the dependent variable. The former Surprise Effect is being replaced by the so-called "Δ-Effect" as not only surprise as a basis emotion mediates the effect. Cognition triggers the incongruence.

Most strikingly the presented Guerilla Effect Model III introduces two highly relevant mediators which include recipient's behavior into the model. Hutter and Hoffmann (2011b, 2013) in their approaches only and exclusively limit the Guerilla Effect to advertising success in relation to strength of a message and range of diffusion, yet ignoring that the result of the Surprise Effect (here Δ-Effect) is either a Communication Behavior or the general Purchase-/Consumer Behavior (Snipes, 2012) whereas such behavior does not exclusively refer to directly achieving monetary goals but also includes stopovers like a general perception with long-term goals in range. Furthermore the Communication Behavior directly mediates to the Diffusion Effect whereas such communication can take place intentionally and unintentionally and also to the Purchase Behavior stating that a behavior can be the result of the prior behavior. The Low Cost Effect is being ignored due to the fact that costs are already considered in the dependent variable as one of the two measures that seek optimization. Listing it as a mediator is not reasonable anymore. Diffusion also directly mediates to the Purchase Behavior being the only mediator variable with a direct path to the desired Guerilla Effect. It can only and exclusively occur via the Purchase Behavior. Such adjustment is necessary to adapt to all-embracing views and unchain the constraint perspective. In the original models a highly entertaining stimulus is being potentially mediated to a Guerilla Effect which only measures the strength of an exposure or quantity of recipients reached. Such perspective is rather limiting than comprehensible. In the end an organization's marketing effort aims at increasing sales. Marketing is not designed to simply entertain a recipient. Just like reality, the model has to be target-oriented and specific and therefore the dependent variable is always mediated via Purchase Behavior. Lastly, a path from the Purchase Behavior to Communication Behavior as mediator variable is necessary to

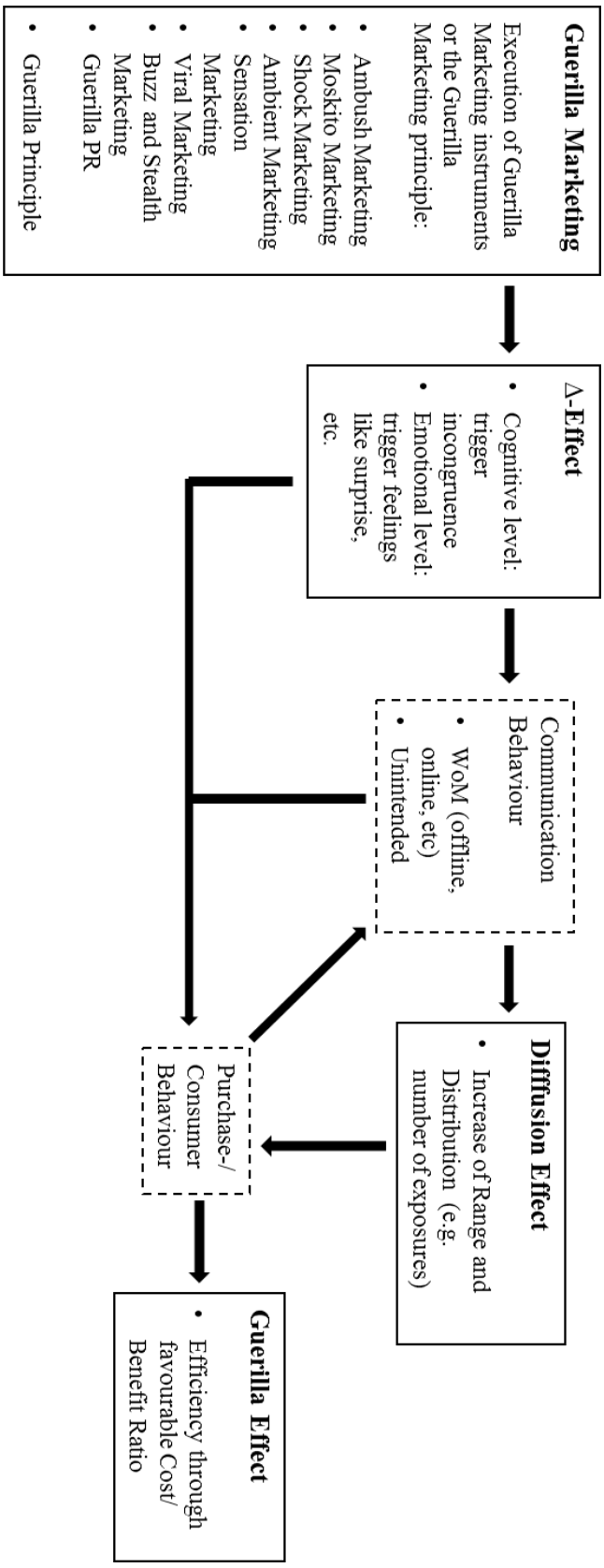

Figure 7: Guerilla Effect Model III; Source: aligned with Hutter and Hoffmann (2011b, 2013).

depict the possible case of further communication and potentially diffusion through a closed sale. Just like before, influencing contexts are being neglected. Finally, the third model introduced should serve as a contribution for further approaches; including various approaches mainly from an overall marketing perspective; yet, also from a psychological perspective trying to explain the Guerilla Effect through the behavior of the recipients or consumers without whom the entire concept would not work.

Most relevantly the models all need testing, e.g. foundation. The upcoming chapter will present the relevant studies trying to (partly) operationalize the model presented. The Guerilla Effect Model III contributed by the author will serve the purpose due to its outreaching design whereas the different instruments only allow limited access to all-embracing approaches. As a result propositions for further research especially regarding modelling Guerilla Marketing will be following.

6. Measuring Guerilla Marketing effectiveness

While scholars claim that unconventional measures of communication can positively influence the credibility of advertising (Dahlén et al., 2009) closer insights are hard to find. Apart from general deficits in measuring advertising efficiency, Krieger (2012, p. 9) states that in Guerilla Marketing a lack of "differentiated observations of the fields and instruments especially in out-of-home advertising" cause diverse issues, while further postulating that "in those few empirical studies and publications many factors of the arrangement of a Guerilla Marketing action (...) remain disregarded". Exemplary in the case of Ambient Marketing challenges arise as "using ambient advertising as a standalone medium of communication is indeed not possible, thus understanding or measuring the impact of effectiveness of an add-on medium like ambient advertising is almost impossible" (Chatterjee, 2011).

Through Algorithms or Emotional Mapping entries have been designed from an expert perspective while surveys aligned with e.g. a Surprise, Diffusion or Low Cost Effect serve as an entry from the consumer perspective (Hutter and Hoffmann, 2013). While monetary and non-monetary input factors to measure the efficiency of advertising can hardly be transferred to Guerilla Marketing, the output measurements (cognitive, affective, behavioral and economical) can serve as a basis for presenting the most relevant studies (Hutter and Hoffmann, 2013).

6.1. Cognitive measurements

Cognitive measures aim at recording to what extent individuals notice (awareness), remember (recall) or recognize a campaign (Kroeber-Riel, 1996). While Wanner (2011, p. 107) states "the main benefit guerilla advertising beholds for its campaigners is brand awareness" further studies also stress great recall effects of Guerilla Marketing in general (Hatch, 2005; Krieger et al., 2012; Hastie and Kumar, 1979). As Ambient Marketing is the instrument obviously showing greatest deviations in the out-of-home advertising, most studies used Ambient Marketing campaigns. Here increased awareness is generated through delimitations from environment and classical out-of-home advertising (Hutter and Hoffmann, 2013) whereas surprising elements are said to be the key for strengthening on downstream emotions (Hutter and Hoffmann, 2013). Following Nufer (2013) such understanding is applicable to any Guerilla Marketing campaign using surprise (e.g. Shock Marketing).

Exemplary in a Recall-Recognition test by Krieger et al. (2012, p. 122) over one third of the surveyed people freely remembered the campaign one day and one week after the exposure whereas strikingly about 20% of the recipients had not witnessed it themselves but heard about it. While measuring awareness in comparison to classical advertising, Guerilla Marketing (e.g. Ambush Marketing) can also be set contraire to sponsors and license holders in order to measure awareness compared to direct competition. In a study of Hutter et al. (2011) the scholars depicted that 74,9% of the spectators assigned three of four ambushers to be official sponsors of the FIFA World Cup 2010. Such designs were frequently used for measuring the increased awareness, recall and recognition for Guerilla Marketing measures whereas increased intention through schema-incongruence as a source for surprise and cognitive stimulation is proved to be the root (Hutter and Hoffmann, 2013).

6.2. Affective measurements

Following Kroeber-Riel (1996) affective measures try to measure whatever processes occur within the consumer (e.g. emotions). General frameworks can be adapted from the field of psychology (Krieger et al., 2012). Hence, reaction tests using Oddball-paradigms measure cognitive processing (Krieger et al., 2012, p. 206) while results vary and prove that not every Guerilla Marketing campaign (e.g. Ambient Marketing) automatically positively influences attention (only one out of two campaigns showed dominating effects). Other studies primarily use surveys and observations to measure image, interest or reputation. A study of 340 real Ambient Marketing campaigns revealed a positive reputation towards Ambient Advertising due to deviation from surrounding stating that Ambient Advertising is more efficient than classical out-of-home advertising (Hutter and Hoffmann, 2013). Such findings are strengthened by Krieger et al. (2012). In regard to reputation the size of an organization plays a crucial role as established brands are expected to continue with similar ways and messages in regard to advertising appropriateness (Wanner, 2011). Furthermore Shock Marketing and advertising that is perceived as being inappropriate can "overpower the rest of the content" whereas again wear-out effects apply (Urwin and Venter, 2014, p. 211).

6.3. Behavioral measurements

While behavioristic approaches frequently revolve around trying to measure purchase behavior in this context and in

regard to the Guerilla Effect Model III (see chapter 5.4.1) the communication behavior also needs consideration (Hutter and Hoffmann, 2013). Apart from basic approaches of identifying motives for engaging in word-of-mouth (Dichter, 1966; Sundaram et al., 1998) studies giving proof for an increased motive in WoM especially via online channels ("eWoM") complement (Hennig-Thurau et al., 2004; Gruber, 2004; Wanner, 2011). Social benefits, economic incentives, the caring for others or self-enhancement are reasons named for the engagement whereas all can be aligned with Guerilla Marketing instruments (Gruber, 2004). While WoM in the context of Guerilla Marketing is predominantly measured by surveys asking for the intention to forward content (Hutter et al., 2011), recall and recognition tests also gave proof for strong WoM effects especially in out-of-home Guerilla Marketing (Krieger et al., 2012). Furthermore Krieger et al. (2012) showed how other social influences (e.g. experiencing a campaign in a group versus individually) further enhanced the effectiveness of a campaign. Recipients in a group showed a significantly greater atmosphere, more intensive interruptions of the current action, activation, surprise and enthusiasm towards the campaign (Krieger et al., 2012). Further the degree of distraction and the level of captivity of the recipients determine the effectiveness (Hutter and Hoffmann, 2013).

6.4. Economic measurements

Closely linked to behavioristic approaches, the economic output measurements gather entail figures like revenue or sales (Hutter and Hoffmann, 2013). Here few approaches deliver close linkages to Guerilla Marketing, as only few studies deliver insights in a direct change in purchase behavior through Guerilla Marketing instruments. (Hutter and Hoffmann, 2013, p. 69) present a proof through scanner panel data for increased sales in a shoe store through the direct effect of POS Guerilla Marketing in front of a store stating an increase in revenue of at least 23%. Surprisingly as the internet allows sophisticated tracking methods and due to the fact that viral campaigns try to close-off sales directly, correlations between a campaign and changes in consumption or sales can be just as challenging as it is for any other Marketing instrument (Langner, 2009).

6.5. Overreaching approaches

Studies recognizing three or more measurements are rare and therefore separately listed. Aligning an undercover observation, survey and the analysis of scanner panel data for an Ambient Marketing campaign Hutter and Hoffmann (2013) give proof for the occurrence of a Surprise Effect (up to 80,4% of the recipients notice the campaign), increased attitude towards the advertising and store aligned with greater eagerness to spread positive WoM and further increased revenue in the store of at least 23,2% during the time of application.

As depicted in the Guerilla Effect Model III, Guerilla Marketing shall follow two intentions through firstly stimulating a purchase intention and secondly stimulating a WoM intention, whereas usually advertising tries to sell over a long term (Adizova et al., 2011). While classical advertising primarily focuses on a product in Guerilla Marketing the instrument plays a relevant role in stimulating WoM and therefore an intention gap can occur between both anticipations. Brand reputation and perceived appropriateness is negatively associated with such intention gap (Adizova et al., 2011). Such findings are being underlined by Hutter and Hoffmann's Sushi-Study (2013) showing that, through exaggeration and antipathy, even though a Guerilla Marketing campaign can evoke much stronger WoM intentions towards the campaign (61,3 to 74,2% increase) than a classical poster, the purchase intentions can be rather lower than with a poster.

6.6. Critique

Recalling the Guerilla Effect Model III introduced in chapter 5.4.1 one can see that only few, if any approaches have been made to operationalize such a model as a whole. The execution of Guerilla Marketing instruments as independent variable bears space of using a great variety of study approaches whereas the more moderator variables involved, the less sophisticated the tests are. While scholars have given proof that Guerilla Marketing instruments can increase attention, surprise and trigger schema-incongruence downstream effects become spongy. Direct purchase behavior or consumer behavior in general is still mostly gathered through asking for the intention to act, but just as hard to measure as it is for any other marketing instrument. In regard to communication, the same rules apply whereas the relevance is even greater for Guerilla Marketing often strongly relying on multiplication via a Diffusion Effect. Just as hard as it is for a practitioner to control such diffusion, it is almost impossible to accurately measure the effects of it. Engaging in WoM, eWoM or anonymous forwarding of content is hard to grasp. Causal relationships even further downstream cannot be measured accurately whilst especially how the effect of purchase behavior on communication behavior generated through a Diffusion Effect closing the cycle in the model is untraceable.

7. Conclusion and recall

Recalling the research questions from chapter 1.2 one can see that ambiguity already arises with the second question. Firstly fitting Guerilla Marketing into the marketing mix as a concept applicable to any marketing tool not limiting it to boundaries or rules is quite more comprehensible than trying to circumvent it against other instruments (e.g. above- or below-the-line, out-of-home, etc.). The diversity of the concept does not allow for restrictions in regard to the field applicability and therefore the proposed bilateral perspective introduced in chapter 2.5 is the least one has to assign to the concept. Circumvention, as asked for in question three, is even harder and due to the amount of instruments almost

impossible. As long as no borders have been cut, vast assignments of instruments will continue to take place in practice and science. Therefore, and in regard to the subsequent presentation of a Surprise and Diffusion Effect where single instruments primarily act as facilitators, the instruments presented in chapter 4 need to be rethought.

Guerilla PR is practically a multiplier of other instruments with little controllability and strong dependence on other instruments. Viral Marketing, also a multiplier, could potentially be considered the maturing of marketing evolution (e.g. WoM) due to technological progress (just like the early days of television or radio advertising) and therefore shall not automatically be assigned to a Guerilla concept. Apart from modern channels for WoM, the “enforced” or “triggered” WoM through Buzz or Stealth Marketing is not any different from classical marketing approaches aiming at positive WoM through customer satisfaction whereas now other forms of intrinsic and extrinsic motivation are being used (Kotler et al., 2007). While Moskito Marketing as a concept is little difference from the everyday effort to identify and make use of competitor’s weaknesses and considering only the ten instruments presented in this thesis, one would be left with only three. Ambush Marketing again is the reaction to increased costs and somewhat a stand-alone image (Nufer, 2013). Interestingly those three are those that are associated with primarily evoking a strong Surprise Effect. Therefore obtaining the perspective of a much narrower field of Guerilla Marketing suspending any other instrument and only focusing on those that really try to use frameless and unconditional measures to maximize attention and surprise would strongly facilitate further practical and scientific approaches and is a justifiable perspective. If the idea of maximizing attention and surprise was transferred to any other instrument or content, one could speak of the application of a Guerilla principle.

Nevertheless and recalling research question number three, "How can Guerilla Marketing be defined regarding the conceptual classification and delimitation?” the author already delivered an extensive derivation and final presentation at the end of chapter three. Such extensiveness would exemplary decline with further limitations as described above.

Modelling Guerilla Marketing has extensively been tackled in chapter 5 whilst an enhanced model has been presented by the author. Considering the lack of empirical research and testing on present models, primarily the downstream variables, no ultimately acknowledgeable model could have been developed, but rather a new kickoff was supposed to be made. Empirically founded models are not identifiable and could therefore neither serve as a basis for further approaches nor conclude research question number six.

The current ambiguity of the field of Guerilla Marketing still only allows scholars to walk on a strongly limited common ground. Besides constantly trying to enlarge the ground for further contributions and interpretations, scholars should possibly narrow down that ground in regard to then being able to have the means to explore it entirely.

Apart from the critical recall the following chapter will briefly present the derived research opportunities for further investigations.

8. Propositions for further research

Firstly, aligned with finalizing definitions Guerilla Marketing should be considered from a more pluralistic perspective granting greater importance to the individual instruments. The diversity of the instruments does not allow for any other handling. Following a circumvention of concept terminology and acknowledged definitions must be the result as a basis.

Sophisticated models for Guerilla Marketing as a whole (if possible) but at least for the given instruments need to be developed. Present and future studies need to be aligned in order to catch up on the scientific shortfall and develop guidelines for practice. As a practical field the controllability (e.g. time-measures) of Guerilla Marketing primarily in regard to a second publicity seek further attention. In line certain instruments need to be excluded or integrated in order to finally round off the concept and be able to have a frame for further investigations.

Psychological and especially behavioristic contributions need to gain further attention in evaluating all influential factors primarily in regard to modelling Guerilla Marketing. Pluralistic consideration is vital again for delivering all-embracing results.

Insights on downstream variables such as the Diffusion Effect will be crucial to fully explore in order to draw clear lines and trace all moderators influencing a Guerilla Marketing Effect.

Lastly, tradeoffs between positive and negative influences and effects have to be identified amongst threshold values for operation and science. Culture, norms and values and other moderators have mostly been ignored so far and need to deserve inclusion into the discussion. Wear-out effects (e.g. Shock Marketing and fading Surprise Effects) have to be considered in order to deliver up-to-date findings.

References

References

Adizova, N., Kadirov, U., and Kadirov, D. Effectiveness of Non-traditional Advertising: Intention Gap and Its Antecedents. In *ANZAM*, volume 383, pages 1–9, 2011.

Ahuja, R., Michels, T., Walker, M., and Weissbuch, M. Teen perceptions of disclosure in buzz marketing. *Journal of Consumer Marketing*, 24(3): 151–159, 2007.

Alden, D., Mukherjee, A., and Hoyer, W. The effects of incongruity, surprise and positive moderators on perceived humor in television advertising. *Journal of Advertising*, 29(2):1–16, 2000.

Ansoff, H. *Corporate strategy: an analytic approach business policy for growth and expansion*. New York, NY: McGraw-Hill, 1965.

Arora, H. Word of mouth in the world of marketing. *Journal of Marketing Management*, 6(4):51–65, 2007.

Ay, C., Aytekin, P., and Nardali, S. Guerrilla marketing communication tools and ethical problems in guerilla advertising. *American Journal of Economics and Business Administration*, 2(3):280–286, 2010.

Bainbridge, J. and Curtis, J. On the ambient trail. *Marketing*, 11:35–39, 1998.

Baltes, G. and Leibing, I. Guerrilla marketing for information services. *New Library World*, 109(1/2):46–55, 2008.

Bigat, E. Guerilla Advertisement and Marketing. *Procedia – Social and Behavioral Sciences*, 51:280–286, 2012.

Böcker, F. and Helm, R. *Marketing*. Lucius & Lucius, Uni-Taschenbücher, 919. Stuttgart, 7 edition, 2003.

Bruhn, M. *Marketing Grundlagen für Studium und Praxis*. GWV Fachverlage GmbH, Wiesbaden, 8 edition, 2007.

Burton, N. and Chadwick, S. Ambush marketing in sport: An analysis of sponsorship protection means and counter-ambush measures. *Journal of Sponsorship*, 2(4):303–315, 2009.

Castronovo, C. and Huang, L. Social Media in an Alternative Marketing Communication Model. *Journal of Marketing Development and Competitiveness*, 6:117–131, 2012.

Chatterjee, S. A study on Ambient Advertising: Marketing Novelty Going Bust. *Research Journal of Social Science and Management*, 1(2):86–91, 2011.

Croft, M. Joking aside. *Marketing Week*, 21(15):41–44, 1998.

Dahlén, M., Granlund, A., and Grenros, M. The consumer-perceived value of non-traditional media: Effects of Brand reputation, appropriateness and expense. *Journal of Consumer Marketing*, 26(3):155–163, 2009.

Dichter, E. How Word-of-Mouth Advertising Works. *Harvard Business Review*, 44(6):147–160, 1966.

Edell, J. and Burke, M. The power of feeling in understanding advertising effects. *Journal of Consumer Review*, 14(12):421–433, 1987.

Egan, J. *Marketing communications*. Thomson, London, 2007.

Fischer, C. *Macht Schlagzeilen! 1000 PR-Ideen, um Kunden und Journalisten für Ihr Unternehmen zu gewinnen*. Gabal-Verlag, Offenbach, 2009.

Gawlik, C., Löffelholz, M., and Redlefsen, J. Modelle der Werbewirkungsforschung am Beispiel der Handywerbung. In *Interkulturelles in der Technik- und Wirtschaftskommunikation*, TU Ilmenau, 2006.

Gelbrich, K., Wünschmann, S., and Müller, S. *Erfolgsfaktoren des Marketing*. Vahlen, München, 2008.

GfK. Marktforschungsstudie zur Nutzung alternativer Werbeformen im Internet. GfK, 2009. URL `http://webguerillas.com/media/press/downloads/pdf/webguerillas_GfK-Studie_2009.pdf`. Accessed: 08.01.2016.

Grauel, R. Unzählige Begegnungen der dritten Dimension. Brand Eins, 2002.

Gruber, M. Marketing in New Ventures: Theory and Empirical Evidence. *Schmalenbach Business Review*, 56(4):164–199, 2004.

Guevara, E. *Der Partisanenkrieg*. Rixdorfer Verlagsanstalt, Berlin, 1982.

Hastie, R. and Kumar, P. Person Memory: Personality Traits as Organizing Principles in Memory for Behavioirs. *Journal of Personality and Social Psychology*, 37(1):25–38, 1979.

Hatch, C. When should you try guerrilla marketing? *ABA Bank Marketing*, 53, 2005.

Hennig-Thurau, T., Gwinner, K., Walsh, G., and Gremler, D. Electronic word-ofmouth via consumer-option platforms: What motivates consumers to articulate themselves in the internet. *Journal of Interactive Marketing*, 18 (1):38–52, 2004.

Hughes, M. *Buzzmarketing: Get People to Talk About Your Stuff*. Portfolio, New York, 2005.

Hutter, K. Guerilla marketing in practice: An exploration into meanings and motives. In *Proceedings of the 12th International Marketing Trends Conference*, 2013.

Hutter, K. and Hoffmann, S. Guerilla-Marketing- eine nüchterne Betrachtung einer viel diskutierten Werbeform. *International Journal of Marketing*, 50 (2):121–135, 2011a.

Hutter, K. and Hoffmann, S. Guerrilla marketing: The nature of the concept and propositions for further research. *Asian Journal of Marketing*, 5(2): 39–54, 2011b.

Hutter, K. and Hoffmann, S. *Professionelles Guerilla-Marketing: Grundlage – Instrumente – Controlling*. Springler Gabler, Wiesbaden, 2013.

Hutter, K. and Mai, R. Effective Incentives for Buzz Marketing: How Moral Concern Moderates the Willingness to Engage as Buzz Agents. *Proceedings of the 42nd Annual Conference of the European Marketing Academy (EMAC)*, 2013.

Hutter, K., Schwarz, U., and Heide, N. Ambush Marketing bei Sportgroßveranstaltungen: Eine empirische Analyse am Beispiel der FIFA Fußball-Weltmeisterschaft 2010. In *Dresdner Beiträge zur Betriebswirtschaftslehre*, volume 160(11), Dresden, 2011.

Hutter, K. and Hoffmann, S. Surprise, surprise. ambient media as promotion tool for retailers. *Journal of Retailing*, 90(1):93–110, 2014.

Jacoby, J. The Emerging Behavioral Process Technology in Consumer Decision-Making Research. *Association for Consumer Research*, 4(1):263–265, 1977.

Javed, M. and Zeb, H. Good shock or bad shock: what impact shock advertisements are creating on the mind of viewers. In *Marketing & Programme Leader International Business: Annual Conference on Innovations in Business & Management [London, 2011, January 26-27]*, pages 1–12, London, University of East London, 2011.

Kaikati, A. and Kaikati, J. Stealth marketing: How to reach consumers surreptitiously. *California Management Review*, 46(4):6–22, 2004.

Keller, E. and Berry, J. *The Influentials: One American in Then Tells the Other Nine How to Vote, Where to Eat, and What to Buy*. Free Press, 2003.

Kotler, P. *Marketing Management*. Prentice Hall, New Jersey, 11 edition, 2002.

Kotler, P., Keller, K., and Bliemel, F. *Marketing-Management: Strategien für wertschaffendes Handeln*. Pearson, Hallbergmoos, 12 edition, 2007.

Krieger, K. H., Esch, F.-R., Osinsky, R., and Henning, J. Die Aktivierungskraft von Guerilla Produktinszenierungen. Ein Vergleich von Guerilla Marketing und klassischer Plakatwerbung mittels Aktivierungsindikatoren im EEG. *Marketing ZFP*, 34(3):196–212, 2012.

Krieger, K. *Guerilla Marketing: Alternative Werbeformen als Techniken der Produktinszenierung*. Springer Gabler, Wiesbaden, 2012.

Kroeber-Riel, W. *Bildkommunikation. Imagerystrategien für die Werbung*. Vahlen, München, 1996.

Langner, S. *Viral Marketing: Wie Sie Mundpropaganda gezielt auslösen und Gewinn bringend nutzen*. Gabler, Wiesbaden, 2 edition, 2007.

Langner, S. *Virales Marketing. Wie Sie Mundpropaganda gezielt auslösen und Gewinn bringend nutzen*. Gabler, Wiesbaden, 3 edition, 2009.

Leonardi, J. *Viral Marketing im E-Business*. Diplomica Verlag, Hermannstal, 2008.

Levinson, J. *Guerilla-marketing – Easy and unexpensive strategies for making big profit from small business*. Mariner Books, New York, 1984.

Levinson, J. *Guerilla Marketing. Offensives Werben und Verkaufen für kleinere Unternehmen*. Campus, Frankfurt/Main, 1998.

Levinson, J. *Guerilla-Marketing des 21. Jahrhunderts: Clever werben mit jedem Budget*. Campus, Frankfurt/Main, 2008.

Levinson, J. and Godin, S. *Das Guerilla Marketing des 21. Jahrhunderts - Clever werben mit jedem Budget*. Campus, Frankfurt/Main, 1996.

Lindsley, D. Emotion. In *Handbook of Experimental Psychology*, page 473–516. Wiley, New York, 1951.

Little, T., Card, N., Bovaird, J., Preacher, K., and Crandall, C. Structural equation modeling of mediation and moderation with contextual factors. In *Modelling Contextual Effects in Longitudial Studies*, pages 207–230. Lawrence Erlbaum Associates, Mahwah, 2007.

Luxton, S. and Drummond, L. What is this thing called 'Ambient Advertising'? In *Proceedings of ANZMAC 2000. Visionary Marketing for the 21th Century: Facing the challenge evaluations*, pages 734–738, Griffith University, Queensland, Australia, 2000.

Mandler, G. The structure of value: Accounting for taste. In *Affect and cognition*, page 3–36. Lawrence Erlbaum, New Jersey, 1982.

Meenaghan, T. Point of view: Ambush marketing: Immoral or imaginative practice? *Journal of Advertising Research*, 34(3):77–88, 1994.

Meffert, H., Burmann, C., and Kirchgeorg, M. *Marketing: Grundlagen marktorientierter Unternehmensführung. Konzepte – Instrumente – Praxisbeispiele*. Gabler, Wiesbaden, 11 edition, 2011.

Menon, V. and Uddin, L. Saliency, switching attention and control: a network model of insula function. *Brain Structure Function*, 214:655–667, 2010.

Meyer, W.-U., Niepel, M., Rudolph, U., and Schützwohl, A. An experimental analysis of surprise. *Cognition & Emotion*, 5(4):295–311, 1991.

Meyer, W.-U., Reisenzein, R., and Schützwohl, A. Toward a process analysis of emotions: The case of surprise. *Motivation and Emotion*, 21(3): 251–274, 1997.

Mohr, I. Buzz marketing for movies. *Business Horizons*, 50(5):395–403, 2007.

Moor, E. Branded spaces-the scope of 'New Marketing'. *Journal of Consumer Culture*, 3(1):39–60, 2003.

Mughari, A. Analysis of guerilla and traditional marketing integration in improving the productivity of organizational marketing in enterprises in Irak: A case study of Kaveh Industrial Estate in Iran. *African Journal of Business Management*, 5:944–948, 2011.

Niepel, M., Schützwohl, A., Rudolph, U., and Meyer, W.-U. Temporal characteristics of the surprise reaction induced by schema-discrepant visual and auditory events. *Cognition & Emotion*, 8(5):433–452, 1994.

Nufer, G. Guerilla Marketing - Innovative or Parasitic Marketing? *Modern Economy*, 4:1–6, 2013.

Nufer, G. and Bender, M. Guerilla Marketing: Reutlinger Diskussionsbeiträge zu Marketing & Management. *Hochschule Reutlingen*, 2008.

Patalas, T. *Das professionelle 1x1: "Guerilla-Marketing - Ideen schlagen Budget"*. Cornelsen, Berlin, 2006.

Porter, M. *The Competitve Advantage: Creating and Sustaining Superior Performance*. Free Press, New York, 1985.

Puttenat, D. *Praxishandbuch Presse- und Öffentlichkeitsarbeit: Eine Einführung in professionelle PR und Unternehmenskommunikation*. Gabler, Wiesbaden, 2007.

Ries, A. and Trout, J. *Marketing warfare*. McGraw-Hill, Hamburg, 1986.

Röthlingshöfer, B. *Mundpropaganda-Marketing: Was Unternehme wirklich erfolgreich macht*. Beck Juristischer Verlag, München, 2008.

Roy, A. and Chattopadhyay, S. Stealth marketing as a strategy. *Business Horizons*, 53:69–79, 2010.

Schulte, T. *Guerilla-Marketing für Unternehmertypen – Das Kompendium*. Sternenfels: Wissenschaft & Praxis, 3 edition, 2007.

Schulte, T. and Pradel, M. *Guerilla-Marketing für Unternehmertypen*. Sternenfels: Wissenschaft & Praxis, 2 edition, 2006.

Seguin, B. and O'Reilly, N. The olympic brand, ambush marketing and clutter. *International Journal of Sports Marketing and Sponsorship*, 4(1/2): 62–84, 2008.

Shani, D. and Sandler, D. Ambush marketing. Is confusion to blame for the flickering of the flame? *Psychology & Marketing*, 15(4):367–383, 1998.

Smith, T., Coyle, J., Lightfood, E., and Scott, A. Reconsidering models of influence: The relationship between consumer social networks and word-of-mouth effectiveness. *Journal of Advertising Research*, 47(4):387–397, 2007.

Snipes, A. Buzzing the Traditional Media off through Ambient Advertising. In *Savannah State University*, 2012.

Solomon, M., Marshall, G., and Stuart, E. *Marketing – real people, real decisions*. Prentice Hall, New Jersey, 7 edition, 2011.

Sundaram, D., Mitra, K., and Webster, C. Word-Of-Mouth Communications: a Motivational Analysis. *Advances in Consumer Research*, 25:527–531, 1998.

Thomas, G. Building the buzz in the hive mind. *Journal of Consumer Behaviour*, 4(1):64–72, 2004.

Tropp, J. *Moderne Marketing-Kommunikation. Systeme – Prozesse – Management*. Verlag für Sozialwissenschaften, Wiesbaden, 2011.

Turk, T., Ewing, M., and Newton, F. Using ambient media to promote HIV/AIDS protective behaviour change. *International Journal of Advertising*, 25(3):333–359, 2006.

Urban, D. and Mayerl, J. Mediator-Effekte in der Regressionsanalyse, 2007. URL http://www.uni-stuttgart.de/soz/soziologie/regression/Mediator-Effekte_v1-3.pdf. Accessed: 08.01.2016.

Urwin, B. and Venter, M. Shock Advertising: Not so Shocking anymore. An Investigation among Generation Y. *Mediterranean Journal of Social Sciences*, 5(21):203–214, 2014.

Varey, R. *Marketing Communications, Principles and Practice*. Routledge, London, 2002.

Walker, R. The hidden (in Plain Sight) Persuaders. The New York Times Magazine, Dezember 5th 2004. 68-75.

Wanner, M. More than the consumer eye can see: Guerrilla advertising from an agency standpoint. *The Elon Journal of Undergraduate Research in Communications*, 2(1):103–109, 2011.

Wehleit, K. *Leitfaden Ambient Media. Grundlagen, Fallstudien und Tipps für erfolgreiches Szenemarketing in jungen Zielgruppen*. Business Village, Göttingen, 2005.

Welling, M. *Guerilla-Marketing in der Marktkommunikation. Eine Systematisierung und kritische Analyse mit Anwendungsbeispielen*. Shaker, Aachen, 2005.

Williams, M. Does Shock advertising still work? Campaign (UK), 16, 11, 2009.

Yüksekbilgili, Z. The use of guerilla marketing in SMEs. *International Journal of Multidisciplinary Research and Review*, 2(2):2–7, 2014.

Zerr, K. Guerilla-Marketing in der Kommunikation – Kennzeichen, Mechanismen und Gefahren. In Kamenz, U., editor, *Applied Marketing*, pages 583–590. Springer, Berlin, 2003.

Junior Management Science 1 (2016) 1-300

Junior Management Science

journal homepage: www.jums.academy

Rückzug auf Raten – Management strategischen Wandels: Eine retrospektive Längsschnittstudie am Beispiel der ChemCo.

Friederike Esther Rhein
Technische Universität Hamburg-Harburg

Abstract

Nicht zuletzt haben die Neuordnung der weltweiten Chemie-Branche infolge von M&A-Aktivitäten und die aktuelle Konsolidierungswelle gezeigt, wie Traditionskonzerne vermehrt wichtige Teile der Unternehmenswurzeln kappen oder in eine „bessere Heimat“ überführen. Angesichts der fortschreitenden Commoditisierung von Kernmärkten und verstärkter Konkurrenz aufstrebender Schwellenländer, trennen sich Unternehmen häufig von weniger rentablen Geschäftssparten, um sich voll und ganz auf lukrative Geschäfte zu konzentrieren und die eigene Zukunft zu sichern. Dass Anpassungen an veränderte Markt- und Wettbewerbsbedingungen und damit verbundene radikale Schnitte häufig auch „zögerlich“ verlaufen können, führt die vorliegende Fallstudie der ChemCo vor Augen. In dem Bemühen verschiedene Wirkungsmechanismen zum spezifischen Zustandekommen eines sukzessiven, über einen Zeitraum von 13 Jahren langfristig vollzogenen „Rückzugs auf Raten“ aufzudecken, werden Rückzugsstrategien innerhalb einer retrospektive Längsschnittstudie unter ihrem Potential zur Implementierung strategischen Wandels untersucht und spezifische Phänomene von Trägheit in den Vordergrund gerückt.

Entgegen ihrer praktischen Relevanz werden Rückzüge als eigenständiges strategisches Werkzeug zur Neuausrichtung in der Managementforschung wenig thematisiert und oft als bloßes Spiegelbild zu M&A-Transaktionen als eine Maßnahme innerhalb Restrukturierungsprogrammen betrachtet. Analysen ist vorrangig eine rein statische Sichtweise zugrunde gelegt. Um tiefergehende Einblicke in das Prozessgeschehen zu gewinnen wurde innerhalb einer retrospektiven Längsschnittstudie auf eine bestimmte Produktgruppe eines deutschen Chemieunternehmens fokussiert und die Geschäftstätigkeit, von den Anfängen im Familienbetrieb bis zum Aufgang im internationalen Großkonzern, untersucht. Die Analyse war darauf gerichtet ein genaueres Bild über spezifische Merkmale und Einflussfaktoren von Rückzugsprozessen zu erlangen und einen Beitrag zu einem besseren Verständnis von Wandlungsbarrieren in der Strategieimplementierung zu leisten.

Basierend auf einer Reihe von problemzentrierten Interviews sowie vielfältigen Dokumentationen und Archivdaten unterschiedlicher Bestände wurden die forschungsleitenden Fragestellungen explorativ untersucht. Unter Zuhilfenahme von Techniken der Grounded Theory werden Formen und Ursachen augenscheinlicher Beharrungstendenzen aus verschiedenen Positionen der Organisationstheorie analysiert. Im Zusammenwirken der Wandlungsbarrieren können verriegelnde Prozesse und dominante Entscheidungsmuster im Rückzugsverlauf aufgedeckt und eine eingetretene „Lock-In-Situation“ ermittelt werden. Erst unter spezifischen Bedingungskonstellationen konnten nachgewiesene Rigiditäten erfolgreich durchbrochen und ein letztlich vollständiger Ausstieg aus schrumpfenden Kerngeschäften langfristig vollzogen werden. Unterschiedliche Konzeptionen der Organisationstheorie, insbesondere Arbeiten zur Pfadabhängigkeit, liefern hierzu wichtige Erklärungsansätze. Damit wird nicht der Anspruch einer Generalisierbarkeit verfolgt, sondern ein Einblick in die Vielschichtigkeit und Multikontext-Probleme des Feldes strategischer Wandelprozesse angestrebt.

Die Forschungsergebnisse erhellen grundlegend das Verständnis für verschiedene Wirkaspekte, welche im Rahmen der Wandlungsvorhaben begrenzenden Charakter besitzen und im Fallgeschehen zu langfristigen Ineffizienzen führten. In Verbindung mit analysierten Verknüpfungen zwischen strategischen Prozessen und Veränderungen des organisationalen Kontext, wird hierbei die Ambivalenz beeinflussender Faktoren, die einerseits als Treiber des Wandels erscheinen, sich gleichsam aber auch in Trägheitstendenzen äußern und zu Pfadabhängigkeiten führen können verdeutlicht. Es wird die Abkehr einer isolierten Betrachtung von Strategie- und Wandelprozessen und Integration einer dynamischen Dimension zum besseren Verständnis von Strategieprozessen betont, wie in der Strategieprozessforschung gegenwärtig verstärkt propagiert.

Keywords: Strategischer Wandel, Rückzugsstrategien, Corporate Restructuring, Organisationale Trägheit, Pfadabhängigkeit

DOI: http://dx.doi.org/10.5282/jums/v1i1pp60-99

1. Einleitung

Das einleitende Kapitel stellt zunächst die empirische Ausgangslage dar, worauf der theoretische Forschungsbedarf aufgebaut und die Forschungsfrage abgeleitet wird. Es folgen ein Überblick zum Forschungsansatz und der zusammenfassende Ausblick auf das Forschungsvorhaben. Die Einleitung schließt mit einer chronologischen Übersicht zum Aufbau dieser Masterarbeit.

1.1. Ausgangslage

Die Erfindung von Acrylglas[1] und Produktentwicklung zu einem breiten Portfolio für vielfältige Anwendungen schufen eine wesentliche Grundlage für den wirtschaftlichen Erfolg der PlasticCo und ebneten neue Wege in die Zukunft.

Am 6. September 1907 wurde die Offene Handelsgemeinschaft PlastiCo zum ursprünglichen Zweck der „Ausbeutung einer von dem Teilhaber (...) erfundenen Beize für Gerbereizwecke“ und des „Handels mit chemischen Produkten“ gegründet (PlastiCo OHG, 1907). Seither hat das Unternehmen in seiner rund 100-jährigen Firmengeschichte einen enormen organisationalen Wandel erlebt und ist heute Teil eines der weltweit führenden Unternehmen der „Spezialchemie“[2] – der ChemCo.

Seit dem Siegeszug des in 1933 markengeschützten „unzerbrechlichen“ PMMA haben sich auch die darauf basierenden Anwendungen von Produkten aus PMMA vielfältig verändert. Weite zivile Verwendung fand das Material im Sanitärbereich, als organischer Werkstoff zur Fertigung von Dusch- und Badewannen. Der Glanz der Aufbaujahre des Sanitärmarktes in den 70er Jahren und Hype auf Acrylglas-Designmodelle wurde jedoch im Verlauf der Marktreifung und Entwicklung zum Commodity-Geschäft, sukzessive von sinkenden Umsätzen, bei Verlust von Marktanteilen an neue Wettbewerber und kontinuierlichem Preisverfall, abgelöst (GmbH, 2007; ChemCo, 2014b).

Der defizitären Geschäftslage entzog sich auch die interne Mitarbeiterschaft nicht. Rückblickend betrachtet wurden seit jeher immer wieder kritische Stimmen laut, welche die Ertragsschwäche des Sanitärgeschäfts benannten[3]:

> „Sanitär war immer die super rote Laterne.“ (#62)

> „Es war klar, glücklich werden wir mit dem Produkt nicht.“ (#210)

> „Man hat den Abwärtstrend gesehen und immer wieder gut geredet, eine Todesspirale.“ (#233)

Nach der ersten Firmenübernahme der PlastiCo durch einen chemischen Großkonzern 1989 und auf die Ertragskrise[4] im Verlauf der 1990er Jahre folgte im Rahmen einer umfassenden Unternehmensrestrukturierung[5] schließlich eine Entscheidung zur Verminderung der Aktivitäten in dem als unprofitabel erwiesenen Sanitärgeschäft. Ein geplanter Rückzug wurde dabei zuerst teilweise umgesetzt, die vollständige Aufgabe der Sanitär-Produktpalette und ein endgültiger Marktaustritt erfolgten allerdings erst 13 Jahre später. Das lange Festhalten an einem verlustreichen Geschäft und der tradierten strategischen Ausrichtung ist aus rational ökonomischer Perspektive kaum zu erklären. Vor diesem Hintergrund werden in der strategischen Managementforschung u. a. strukturelle und emotionale Ursachen als Beharrungskräfte auf individueller wie systemischer Ebene thematisiert.

Die vorliegende Masterarbeit soll einen Beitrag zu einem besseren Verständnis von Rückzugsstrategien im Kontext von Prozessen strategischen Wandels und der Erforschung einer augenscheinlich rational nicht zu erklärenden Entscheidungspersistenz wirtschaftlich ineffizienter Strategien aus einer Perspektive des strategischen Managements leisten.

1.2. Problemstellung und Forschungsziele

Veränderung von Industrie und Geschäftsmodellen kraft dynamischer[6] Unternehmensumwelt impliziert unter anderem auch, dass selbst marktführende Unternehmen das Engagement ihrer produkt-/marktbezogenen Aktivitäten im Verlauf der Geschäftstätigkeit vermindern, Produkte aufgeben und sich letztlich ganz aus Märkten zurückziehen können.

Ein vielfach zu beobachtendes Phänomen in diesem Zusammenhang sind die zögerlichen Anpassungsprozesse zur Entwicklung und Implementierung solcher Rückzugsstrategien. Die fortwährenden Veränderungen innerhalb von Unternehmen sind dabei in einem Spannungsfeld von Kräften zum Wandel und Gegenkräften der Beharrung zu sehen. Perioden langanhaltenden Erfolgs sowie der langfristige Auf- und Ausbau spezifischer Kernkompetenzen und Ressourcen bergen die Gefahr dem Wandel entgegenwirkende Beharrungs-

[1] Synthetisch hergestellter thermoplastischer Kunstoff Polymethylmethacrylat (Kurzzeichen PMMA). Die Bezeichnungen werden im Folgenden synonym verwendet.

[2] Vor dem Hintergrund einer nicht einheitlichen Begriffsdefinition, werden hierunter chemische Produkte subsummiert, die für spezielle Anwendungen, häufig kundenspezifisch, entwickelt werden, wobei die Wirkung für das Endprodukt und weniger die chemische Beschaffenheit im Vordergrund steht. Häufig kommen diskontinuierliche Produktionsverfahren zum Einsatz, die Herstellkosten sind gegenüber den in Massenproduktion hergestellten Grundchemikalien sehr hoch. Im Gegenzug können vergleichsweise hohe Preise erzielt werden.

[3] Retrospektive Aussagen aus den Experteninterviews damaliger interner Akteure.

[4] Zur Beschreibung des Verlaufs von Unternehmenskrisen können Unternehmenskrisen nach Kategorien, z. B. anhand des Grades der Existenzbedrohung, typologisiert werden: Strategiekrisen beziehen sich auf den Verlust von Erfolgspotentialen, Ergebniskrisen sind durch das Nichterreichen von Gewinn- und Rentabilitätszielen, die Liquiditätskrise durch Gefährdung der Zahlungsfähigkeit charakterisiert (vgl. z. B. Müller, 1986 sowie Krystek und Moldenhauer, 2007; Bamberger und Wrona, 2012b:47-56, Krüger, 2002:59ff.).

[5] Unter dem Begriff der Unternehmensrestrukturierung (auch Reorganisation) können Veränderungen von Prozessen, Systemen und Strukturen sowie materiellen Potentialen subsummiert werden, die auf eine Anpassung der Unternehmensstruktur und –aktivitäten an veränderte Marktgegebenheiten abzielen. In der Literatur werden verschiedene Formen und Elemente unterschieden (vgl. hierzu z. B. die Klassifizierung von Bowman und Singh, 1993 in Online-Appendix 1).

[6] Siehe zur Entwicklung des Dynamikbegriffs und konzeptionellen Spezifizierung u. a. Perich, 1992:91-116.

tendenzen zu entwickeln und in eine Trägheitsfalle zur Aufrechterhaltung des „Status quo“ zu geraten.[7]

In solchen Situationen sind die Ursachen dafür, dass Unternehmen trotz Anpassungserfordernissen nicht in der Lage sind, sich zu wandeln oder grundlegende Veränderung konterkariert wird, oft unklar.

Im Zusammenhang von Wandel und Veränderung wird in der Wissenschaft über „organisationale Trägheit“ oder „strategische Rigidität“ diskutiert (vgl. u. a. Huff et al., 1992:56, Hannan und Freeman, 1984, 1989).

Diese Phänomene von Beharrungstendenzen und wirksamen Wandelbarrieren bei der Realisierung von Veränderungsprozessen - wie im Falle des Rückzugs aus dem Sanitärgeschäft der damaligen PlastiCo - sowie die Frage nach den Ursachen von Trägheit und Widerstand stehen im Zentrum der vorliegenden Masterarbeit.

Die Thematisierung von Rückzugsstrategien zur Implementierung nachhaltigen Wandels ist mit Blick auf die aktuelle Forschung zur strategischen Führung besonders relevant, da Rückzüge angesichts einer dominierend wachstumsorientierten Ausrichtung einen vernachlässigten Gegenstandsbereich darstellen und häufig mit Scheitern bzw. der Korrektur zuvor getroffener Entscheidungen gleichgesetzt werden. In der Mehrzahl diesbezüglich vorgelegter empirischer Studien werden nur punktuelle Aspekte im Rückzugsprozess fokussiert und oft eine statische Sichtweise zum Zustandekommen von Rückzugsstrategien zugrunde gelegt. Der strategische Wandlungsprozess zwischen den „Snapshots“ vor und nach dem Rückzug wird nicht näher untersucht, sodass sich im Sinne einer explorativen und illustrativen Funktion in der qualitativen Forschung hinsichtlich der Realisierung von Rückzugsstrategien und ihren Bezug zu neuralgischen Kontextmerkmalen eine vielversprechende Oberfläche der Generierung von Erkenntnissen bietet.

Durch den explorativen Charakter dieser Arbeit soll ein genaueres Bild über Strategie- und Wandlungsprozesse sowie ihrer Interdependenz im Kontext der Unternehmensentwicklung und externen Umwelt erlangt und ein Beitrag zur Erfassung und interpretativen Erklärung des Trägheits-Phänomens innerhalb von Prozessen strategischen Wandels geleistet werden.

Im Rahmen einer retrospektiven Längsschnittstudie am Beispiel der Geschäftsaktivitäten der ursprünglichen PlastiCo bestehen konkrete Forschungsziele darin:

- *Deskriptiv die Art und den inhaltlichen Verlauf von strategischen Wandelprozessen darzustellen und hierbei spezifische Dimensionen aus handlungsorientierter Perspektive unter Rückgriff theoretischer Konzeptionen zu analysieren,*
- *explikativ aufzuzeigen, ob Organisationen beständig gegen Wandel zu sein neigen und in welcher Art langfristig versucht wird an verlustreichen Strategien festzuhalten sowie*
- *der Verhaltensweise zugrunde liegende Ursachen herauszustellen, die prägend für das Spannungsfeld von Trägheit und Veränderung sind.*

Mit der empirischen Untersuchung soll nicht der Anspruch einer Generalisierbarkeit erhoben, sondern ein Einblick in die Vielschichtigkeit und spezifische Problembereiche des Feldes organisationalen Wandels sowie, in einer integrierten Betrachtung von Strategie- und Wandelprozessen, Erkenntnisse zum praktischen Phänomen von Trägheitskräften in Organisationen und deren Kontextbezug gewonnen werden.

Als Ergebnis der Fallstudienanalyse soll eine deskriptive Darstellung verschiedener Ausprägungsformen von Trägheit beachteter Wandelprozesse erfolgen und zur Erklärung ein Verweisungszusammenhang zu theoretischen Konzeptionen der Organisationstheorie hergestellt werden.

Mit diesem Untersuchungsschwerpunkt soll ferner die im Gesamtzusammenhang der Strategieprozessforschung unterrepräsentierte Problematik der Strategieimplementierung[8] - im Vergleich zu Aspekten der Strategieformulierung - in den Vordergrund gerückt bzw. auf die Beziehung zwischen Strategieentwicklung- und Implementierung weiter Bezug genommen werden. Hintergründe zu spezifischen Problembereichen der Realisierung von Rückzugsstrategien sollen Manager aus der Unternehmenspraxis für eine zielgerichtete Gestaltung von Veränderungsprozessen im Sinne eines erfolgreichen und effizienten „Change Managements“ (vgl. Bamberger und Wrona, 2012a: 513-522) sensibilisieren.

1.3. Methodik und Aufbau der Arbeit

Die skizzierten Forschungsfragen werden auf Grundlage einer systematischen Fallstudienanalyse[9], basierend auf einer Reihe von problemzentrierten Interviews sowie vielfältigen Dokumentationen und Archivdaten unterschiedlicher Bestände, explorativ untersucht. Durch die große Nähe zum Datenmaterial, die vor allem mit der Durchführung persönlicher Interviews gegeben ist, können mit der qualitativen Fallstudienforschung komplexe Sachverhalte tiefgreifend analysiert und Aufschlüsse über kausale Wirkzusammenhänge gegeben werden (vgl. Bortz und Döring, 2006; Lamnek, 2005:89,109). Diese Charakteristika sind für die vorliegende Arbeit von hoher Relevanz, weil die im Fokus stehenden Phänomene einen vielschichtigen Problembereich darstellen, und zu betrachtende Prozesse infolge eines stark subjektiv geprägten Einflusses individueller wie kollektiver

[7] Vgl. exemplarisch Tripsas und Gavetti, 2000; Burgelman, 2002.

[8] Obgleich die Problematik der Implementierung strategischer Pläne - im Sinne eines organisationalen Wandels zur Anpassung an die jeweilige Strategie - schon frühzeitig, zurückgehend auf die Arbeit von Chandler (1962) aus dem Jahre 1962, in der Literatur erkannt wurde, wird diesem Forschungsfeld noch immer wenig Aufmerksamkeit gewidmet. Vor allem in der deutschsprachigen Literatur werden vorrangig Aspekte der Strategieformulierung thematisiert (vgl. Zahn, 1999: 3, Hilker, 1993: 2).

[9] Unter Berücksichtigung eines nicht homogenen Begriffsverständnis zum Terminus „Fallstudie“ orientiert sich diese Arbeit an der Beschreibung der Fallstudienanalyse als wissenschaftliche Forschungsmethode nach Wrona, 2005.

Wahrnehmung und Verarbeitung sowie politischer bzw. sozialer Aspekte, mit eher „weichen“ Attributen verbunden sind.

Zur Ergründung des ineffizienten Entscheidungsmusters in Wandlungsprozessen wird zunächst eine Inhaltsperspektive eingenommen und zunächst der organisationale und strategische Entwicklungsverlauf über die Betrachtung relevanter Einfluss- bzw. Kontextgrößen im Analysezeitraum nachgezeichnet.

Zur fundierten Analyse und Klärung der forschungsleitenden Fragestellung nach dem Zustandekommen eines sukzessiven, über einen Zeitraum von mehr als 10 Jahren langfristig vollzogenen „Rückzugs auf Raten“ und um das Verständnis verschiedener Prozessdimensionen zu schärfen, wird die inhaltsorientierte Sichtweise im Verlauf der Arbeit über eine handlungsorientierte Prozessperspektive, Bezug nehmend auf die Thematik „organisationaler Trägheit“, ergänzt und dem gegenseitigen Verweisungszusammenhang beider Sichtweisen innerhalb der strategischen Unternehmensführung damit Rechnung getragen.[10]

Um den realisierten Marktaustritt hinsichtlich der Ausprägungsform und Ursachen vermuteter Trägheitskräfte näher zu untersuchen, wird dabei als klassischer Ausgangpunkt an den Grundannahmen des deterministisch orientierten Population Ecology-Ansatz, basierend auf den Arbeiten von Hannan/Freeman aus dem Jahre 1977, angeknüpft.

Aspekte aus alternativen theoretischen Erklärungsansätzen wie der volkswirtschaftlich orientierten Industrieökonomik (vgl. u. a. Porter, 2008), kognitiv-interpretativen (vgl. z. B. Walsh, 1995; Mintzberg, 1990; Rajagopalan und Spreitzer, 1996) und politischen Konzeptionen (vgl. u. a. Easton, 1965) sowie Erkenntnissen aus der Pfadtheorie (vgl. u. a. Schreyögg und Sydow, 2003) werden zuzüglich aufgegriffen und deren Erklärungskraft zusammenhängend diskutiert.

Die Arbeit ist in vier Kapitel gegliedert, die sich am einleitend dargestellten Forschungsdesign orientieren. Im nachfolgenden zweiten Kapitel wird in die Thematik von Unternehmen im Spannungsfeld von Veränderung und Trägheit zunächst theoretisch eingeführt. Dazu werden zuerst Grundkonzeptionen der strategischen Unternehmensführung dargestellt und eine definitorische Einordnung zum Strategiebegriff gegeben. Anschließend wird das Feld strategischer Prozesse der Wandels im Mittelpunkt stehen, wobei speziell Rückzugsstrategien zum strategischen Richtungswechsel fokussiert werden. Bezug nehmend auf beeinträchtigende Faktoren des unternehmerischen Wandels und Möglichkeiten zur Steuerung von Veränderungsprozessen wird versucht, verschiedene Perspektiven von Wandlungsprozessen zu einem integrativen Konzept zusammenzuführen, welches das theoretische Grundgerüst für die empirischen Untersuchung bildet. Hierauf folgt in Kapitel 3 die empirische Untersuchung, die den Schwerpunkt der Arbeit darstellt. Diese beginnt mit der Darlegung des methodischen Vorgehens in Bezug auf die forschungsleitende Problemstellung und Beschreibung des Untersuchungsgegenstandes. Basierend auf dem im Theorieteil entwickelten Bezugsrahmen wird daraufhin eine systematische Analyse durchgeführt und dabei beobachtete Phänomene klassifiziert und eingeordnet. Die zuvor genannten Forschungsfragen werden in diesem Schwerpunktkapitel im Rahmen eines zweischrittigen Vorgehens behandelt: Eine Darstellung des historischen Entwicklungsverlaufs und die nachfolgende Trägheits-Analyse. Abschließend werden empirische Befunde theoretisch nachvollzogen und diskutiert. Der Schlussteil in Kapitel 4 rundet die Arbeit mit einem zusammenfassenden Fazit und Implikationen für Theorie und Praxis ab. Abbildung 1 stellt das Vorgehen noch einmal grafisch dar.

2. Theoretische Grundlagen der strategischen Unternehmensführung & Management des Wandels

Innerhalb dieses Kapitels sollen die theoretischen Grundlagen, auf denen die empirische Untersuchung in Kapitel 3 basiert, gelegt werden. Dazu werden in einem ersten Unterkapitel zentrale Grundkonzeptionen der strategischen Führung von Unternehmen behandelt und ein für die Arbeit notwendiges Begriffsverständnis von „Strategien“ gegeben. Hierauf aufbauend widmet sich das zweite Unterkapitel den strategischen Prozessen im Rahmen der strategischen Unternehmensführung. Diese sollen speziell hinsichtlich ihrer dynamischen Dimension, als Prozesse des Wandels, thematisiert und der Versuch unternommen werden, eine eigene integrative Sichtweise von Strategie- und Wandelprozessen zu entwickeln. Dabei werden speziell Rückzugsstrategien als eine Art der Implementierung strategischen Wandels näher beleuchtet und im Zusammenhang mit Fragestellungen zum Zustandekommen von Wandel und einflussnehmenden Faktoren, Phänomene organisationaler Trägheit und des Widerstandes gegen Wandel sowie Ansätze zum Management behandelt. Der Theorieteil endet mit der Vorstellung der eigenen Konzeption zur Zusammenfügung verschiedener Perspektiven auf strategische Prozesse des Wandels in Form eines Bezugsrahmens, der eine breite und differenzierte Analyse vielschichtiger Sachverhalte erlaubt und in einer erkenntnisleitenden Funktion der empirischen Untersuchung zugrunde gelegt wird.

2.1. Grundkonzeptionen der strategischen Unternehmensführung

2.1.1. Perspektiven und Sichtweisen der strategischen Unternehmensführung

Die Analyse von Aspekten der Führung von Organisationen[11] und besonders Unternehmen hat eine lange Tradition innerhalb der Managementforschung, Betriebswirtschaftslehre sowie der Organisationstheorie. Sie hat dabei eine

[10] Siehe zur Leitdifferenz zwischen inhalts- und prozessbezogenen Perspektive der strategischen Unternehmensführung Abschnitt 2.1.1.

[11] Die vorliegende Arbeit folgt einem institutionellen Begriffsverständnis von Organisationen als zielgerichtete soziotechnische Systeme (vgl. z. B. Schreyögg, 2006:4-11, Gomez und Zimmermann, 1992:16, als Überblick zur Definitionsvielfalt auch Vahs, 2007:10ff.).

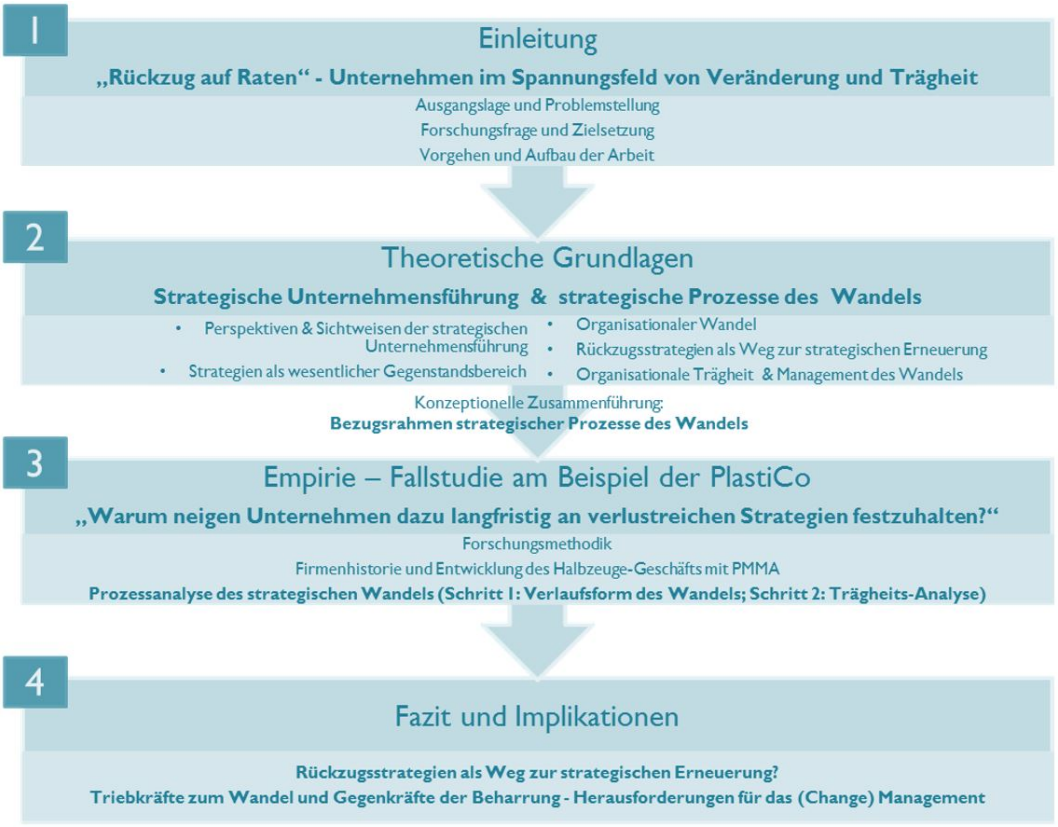

Abbildung 1: Aufbau der Arbeit

Vielzahl theoretischer Perspektiven der strategischen Unternehmensführung hervorgebracht, wobei einige, die für diese Arbeit bedeutsam sind, im Folgenden skizziert werden sollen.

Strategische Unternehmensführung[12] wird hierbei im funktionalen Sinne, als jene Art der zielgerichteten Lenkung, Gestaltung und Entwicklung der Unternehmung verstanden, die sich auf den Aufbau von „Erfolgspotentialen“[13] zur Realisierung zukünftiger Erfolge bezieht. Gegenstandsbereiche bilden typischerweise die Bestimmung von Zielen, Strategien, grundlegendem Ressourceneinsatz und auch Grundsätzen sowohl auf der Ebene der Gesamtunternehmung, wie auch Geschäftsbereiche und -Einheiten. Ebenso sind speziell in Form von Managementsystemen, wie IuK-Systeme, Personalführungssysteme und Planungs- und Kontrollsysteme, Instrumente zur Führung der Unternehmung zu schaffen.

Unter Berücksichtigung, dass mit den einzelnen Gegenstandsbereichen die langfristige[14] Entwicklung der Unternehmung bestimmt wird, kann strategische Unternehmensführung auch auf die Beeinflussung organisationalen Wandels gerichtet sein. Die Anpassung an Veränderungen der Umwelt[15] ist dabei als grundlegendes strategisches Problem zu sehen, womit deutlich wird, dass Unternehmensführung stets auch die Führung des Wandels ist.

Prozesse der strategischen Unternehmensführung sind kontinuierliche Prozesse, wobei die Bildung und Veränderung von Handlungsorientierungen[16], das Treffen von Entscheidungen[17] und Maßnahmen zur Umsetzung interagieren.

In Abgrenzung zur operativen und taktischen Ebene liegen strategischen Problemen schlecht strukturierte Problemdefinitionen zugrunde. Es handelt sich um innovative Probleme in dem Sinne, als dass keine eindeutigen Lösungsalgo-

[12] Die Begriffsverwendung im Kontext der strategischen Unternehmensführung ist sowohl international als auch im deutschsprachigen Raum nicht einheitlich. Dieser Arbeit lehnt an dem Begriffssystem nach Bamberger und Wrona, 2012a:32ff. an.

[13] Der ursprünglich von Gälweiler, 1974 eingeführte Begriff des „Erfolgspotentials“ wird in der Literatur unterschiedlich definiert (siehe etwa Gälweiler, 2005; Nähter, 1993; Wienand, 1982) und sei hier im Sinne von Positionen, welche Vorlaufsgrößen für zukünftige Erfolge bilden, verstanden (vgl. Bamberger und Wrona, 2012a:20).

[14] Die langfristige Perspektive kann hierbei in verschiedener Weise interpretiert werden: Einerseits kann die Wirkung von Entscheidungen sowie andererseits der Bezugszeitraum der Entscheidung zur Entwicklung der Unternehmung langfristig sein (vgl. Bamberger und Wrona, 2012a:13).

[15] Vor dem Hintergrund kontingenztheoretischer Überlegungen des „Strategic Fit“ gilt die Unternehmensumwelt als grundlegender Referenzpunkt zur Notwendigkeit einer Anpassung von Organisationen. Danach wird der Unternehmenserfolg durch die Kohärenz bzw. Stimmigkeit verschiedener interner und externer Faktoren bestimmt (vgl. Venkatraman, 1989; Venkatraman und Camillus, 1984; Miller, 1988; Bea und Haas, 2009: 405).

[16] Als Modelle zukünftigen Verhaltens, die mehrere Entscheidungsepisoden überdauern und handlungsleitende Kraft entfalten (vgl. Bamberger und Wrona, 2012a:102ff.).

[17] Zum Entscheidungs-Begriff vgl. ausführlich z. B. Kirsch, 1988, als Überblick über strategische Entscheidungsprozesse Bamberger und Wrona, 2012a: 457-462.

rithmen bestehen. Sie sind stark wertbeeinflusst und mithin auch als sogenannte „Multi-Kontext-Probleme“ (vgl. Kirsch, 1990:12ff.) zu bezeichnen.

In der Betrachtung des breiten Feldes möglicher Merkmale bzw. Themenkomplexe der strategischen Unternehmensführung hat sich im wissenschaftlichen Diskurs eine Unterscheidung zwischen einer inhaltsorientierten (Strategy Content Research) und prozessorientierten Sichtweise (Strategy Process Research) als Leitdifferenz etabliert, die für die disziplinäre Entwicklung weiterhin maßgeblich prägend ist.[18]

Die Strategy Content Forschung beschäftigt sich primär mit der Beschreibung und Erklärung „strategischer Verhaltensweisen“ und im Kern Strategien selbst als Ergebnis des Strategieprozesses zur Erreichung des dauerhaften Unternehmenserfolgs.

Hingegen werden innerhalb einer prozessbezogenen Perspektive diejenigen Aktivitäten bzw. Ergebnisse innerhalb einer Unternehmung untersucht, über die „strategische Phänomene“ zustande kommen. Im Vordergrund stehen z. B. Fragestellungen nach der Herausbildung von Strategien, ihrer Planbarkeit und Bedeutung maßgeblich beteiligter Akteure im strategischen Prozess. Beide Perspektiven können und müssen aufeinander bezogen werden, wie im Verlauf der vorliegenden Arbeit noch gezeigt wird (Bamberger und Wrona, 2012a: 5ff. und 414f., Müller-Stewens und Lecher, 2001: 11ff., Welge und Al-Laham, 2008:159).

2.1.2. Strategien als Gegenstandsbereich einer strategischen Unternehmensführung

Der Terminus „Strategie“ (von altgriechisch „strategós“ „Feldherr, Kommandant“) meint in seiner ursprünglichen Bedeutung die „Kunst der Heerführung“ (vgl. Kreikebaum et al., 2011:24f.) und bezieht sich auf die Erstellung einer übergreifenden Gesamtkonzeption zur Erreichung langfristiger Ziele. Den Einzug in die Wirtschaftswissenschaft erhielt das Strategiekonzept in den 50er Jahren an der Harvard Business School (Snow und Hambrick, 1980), seitdem wird der Begriff der Strategie in Wissenschaft und Praxis uneinheitlich verwendet und dabei auf unterschiedliche Phänomene bezogen.

Im Rahmen dieser Arbeit soll als Grundlage die Perspektive von (Mintzberg, 1985), welcher Strategien als ein Muster in einem Strom von Entscheidungen bzw. Handlungen betrachtet, eingenommen werden. Mintzberg geht davon aus, dass Strategien sowohl willentlich geplant als auch unwillentlich durch unterschiedliche Entscheidungen innerhalb und außerhalb einer Unternehmung zustande kommen können und unterscheidet intendierte, unrealisierte, deliberate, emergente und realisierte Strategien (Mintzberg, 1977). Intendierte (beabsichtigte) Strategien beinhalten Pläne, die in Form von Beschränkungen und Leitlinien des zukünftigen Verhaltens dienen sollen. Realisierte Strategien ergeben sich ex post als Verhaltensmuster. Intendierte Strategien müssen nicht notwendigerweise realisiert werden (unrealisierte Strategien). Realisierte Strategien können durchaus auf Plänen beruhen (deliberate Strategien), sich alternativ aber auch „emergent“ herausbilden, weil sie z. B. anfangs nicht vorgesehen waren oder zwischenzeitlich wieder verworfen wurden. Strategien können mithin auf Plänen beruhen, sich emergent ergeben oder sich durch das Zusammenspiel von Umwelteinflüssen, unternehmensinternen Kräften und der Führung des Unternehmens formieren (vgl. Mintzberg und Waters, 1985, S. 257 ff., Mintzberg et al., 1998).[19]

Für diese Arbeit sollen weiterhin verschiedene Strategiearten[20] innerhalb eines Strategiesystems unterschieden werden, anhand derer sich strategischer Wandel erfassen lässt. Ausgangpunkt dabei soll eine Klassifizierung von drei Strategieebenen einer rein logischen Hierachie bilden. Eine Betrachtung von Unternehmens-, Geschäfts(feld)- und funktionalen Strategien soll die seit den 80er Jahren prägende Denkweise in strategischen Geschäftsfeldern[21] sowie kennzeichnende Schlüsselbegriffe von Wettberwerbsvorteilen und Kernkompetzenten zur Analyse strategischen Verhaltens zum Ausdruck bringen. Zwischen den drei Arten von Strategien bestehen dabei unmittelbarere Interdependenzbeziehungen. Unternehmensgesamtstrategien beziehen sich auf die Bestimmung globaler Aktivitätsstrukturen, in der Art von Produkt-Markt-Kombinationen bzw. strategischer Geschäftsfelder, sowie die Wahl von Strategien zum Eintritt und ggf. der Aufgabe dieser Aktivitätsfelder. Geschäftsfeldstrategien beeinhalten globale Aktivitätsstrukturen zur Zielerreichung in Bezug auf die einzelnen strategischen Geschäfte und können nach mehreren Kriterien unterschieden werden. Hierbei sollen zwei grundsätzliche Arten anhand der Unterscheidungsmerkmale aufzubauender Wettbewerbsvorteile sowie der intendierten strategischen Stoßrichtung zur Erreichung einer bestimmten Wettbewerbsposition betrachtet werden. Dabei können zum einen Wettbewerbsstrategien zur Erreichung von Wettbewerbsvorteilen[22], welche, dem Konzept der generischen Wettbwerbsstrategien nach Porter folgend, weiterhin in die drei Strategieklassen der Kostenführerschaft, Differenzierung sowie Fokussierung auf Schwerpunkte in

[18] Vgl. für Systematisierungen im deutschen Sprachraum u. a. Al-Laham, 1997:22f., Knyphausen-Aufseß, 1995; Kirsch und Knyphausen-Aufseß, 1993; Habel, 1992; Sydow, 1992:237ff.; im amerikanischen Sprachraum: z. B. Schendel, 1992a,b; Pettigrew, 1992; Harrison, 1990; Huff und Reger, 1987; Schendel und Cool, 1986; Snow, 1986; Fahey und Christensen, 1986.

[19] Aufbauend auf der Unterscheidung von bewussten und emergenten bzw. intendierten und realisierten Strategien entwickelten Mintzberg und Waters das Konzept des „Strategic Learnings“, welches den Lernprozess innerhalb der Strategieentwicklung, basierend auf einem Vergleich der intendierten Strategien mit den tatsächlich realisierten Strategien, sowie eine auf den Erfahrungen aus diesem Lernprozess resultierende Modifizierung zukünftig zu planender Strategien berücksichtigt (vgl. Mintzberg und Waters, 1985:271).

[20] Einen systematischen Überblick über Arten von Strategien anhand unterschiedlicher Kriterien liefern z. B. Kreikebaum et al., 2011:137.

[21] Im Sinne eng miteinander verbundene Produkt-/Markt-Aktivitäten, auch als „ strategische Geschäfte“ bezeichnet (vgl. Bamberger und Wrona, 2012a:111). Dort, wo sich Unternehmen am Kunden und Marktanwendungen ausrichten, sind Geschäftsfelder und Marktsegmente in der Regel identisch. Die Begriffe Geschäftsfeld, Geschäft und Marktsegment werden hier synonym verwendet.

[22] Beziehen sich auf Positionsvorteile, die eine Unternehmung auf einem Markt gegenüber ihren Konkurrenten besitzt (vgl. Bamberger und Wrona, 2012a:121).

Form von Kosten- sowie Differnezierungsvorteilen untergliedert werden können, sowie andererseits Geschäftsentwicklungsstrategien in Bezug auf verfolgte Entwicklungsziele unterschieden werden. Geschäftsentwicklungsstrategien stehen in enger Beziehung zur jeweiligen Marktevolutionsphase und Wettbewerbsposition der Unternehmung.[23] Neben Strategien zur Marktanteilserhöhung sowie Konsolidierungsstrategien bilden für die Problemstellung in dieser Arbeit die Klasse der Rückzugs- bzw. Desinvestitionsstrategien den Bezugspunkt, weswegen diese im Zusammenhang mit der Thematik strategischen Wandels im nachfolgenden Teil eingehender betrachtet werden.

Abschließend sei auf die dritte und letzte Ebene funktionaler Strategien als globale Aktivitätsstrukturen innerhalb verschiedener Funktionsbereiche der Unternehmung, wie bspw. F&E, Produktion, Logistik oder Marketing, hingewiesen (vgl. ausführlich Bamberger und Wrona, 2012a:106ff.).

2.2. Eigene Perspektive zu Wandlungsprozessen

2.2.1. Grundlagen des Wandels

Strategische Prozesse[24], die im Kern u. a. die Entwicklung und Implementierung von Strategien beinhalten, bringen Veränderungen von Merkmalen der Organisation hervor und sind damit eng mit organisationalen Wandel verbunden.

Wandel lässt sich in allgemeiner Weise als Veränderung von Systemmerkmalen in der Zeit[25] definieren. Organisationaler Wandel bezieht sich insofern auf die Änderung der Gesamtheit aller Merkmale von Organisationen im Zeitverlauf (vgl. (Bamberger und Gabele, 1977)). Dieser sehr universale Begriff des organisationalen Wandels kann sich mithin auf vielfältige Objektbereiche beziehen und dabei sehr unterschiedliche Arten von Veränderungsprozessen einschließen, die in der Organisationstheorie und BWL durch eine Vielzahl von Begrifflichkeiten, wie u. a. Restrukturierung, Evolution, Adaption, (Organisations-) Entwicklung, Reorientierung, Erneuerung und Transformation, erfasst wird.

Organisationaler Wandel lässt sich danach klassifizieren, welche Objekte einer Organisation angesprochen werden. Eine Kennzeichnung verschiedener Objektbereiche der Veränderung kann anhand einer Unterscheidung von Veränderungen in der Oberflächen- bzw. Tiefenstruktur der Organisation erfolgen.[26] Im Gegensatz zur von außen sichtbaren bzw. ex post rekonstruierbaren Oberflächenstruktur, bilden jene Merkmale einer Unternehmung die Tiefenstruktur, welche nicht aus einer Außenperspektive beobachtet werden können. Die wechselseitige Beziehung von Oberflächen- und Tiefenstruktur findet in der Organisationslehre im Rahmen sogenannter „Schichtenmodelle“ als Klassifikationsraster ihren Ausdruck. In Analogie zur Struktur einer Zwiebel sind demnach Oberflächenschichten jeweils in hierarchisch tieferliegende Schichten eingebettet, die diesen eigentlich zugrunde liegen. Die einzelnen Schichten drücken danach einen bestimmten Grad der Kopplung und zunehmende Schwierigkeit des Zugangs aus und sind mit verschiedenen Formen bzw. Dimensionen von Wandlungsprozessen verbunden. Hingewiesen sei beispielhaft auf die Modelle von Tushman und Romanelli (1985, 179f.), Perich (1992, 133f.) und Krüger (1994, 358ff.) (vgl. Abbildung 2).

Ferner werden bei der Analyse von Organisationsstrukturen nicht nur vertikale, sondern auch horizontale Verknüpfungen im Rahmen der Interdependenzbeziehungen einzelner Objekte einer „Schicht“ von Phänomenen einer Organisation betrachtet (vgl. Bamberger und Wrona, 2012a:494ff.).

Für eine allgemeine Charakteristik des Untersuchungsobjekts werden in der Literatur verschiedene Arten organisationalen Wandels definiert, wesentliche seien hier anhand gegensätzlicher Begriffspaare skizziert:

In Bezug auf die eingangs erläuterte Kennzeichnung des „Strategischen“ kann grundlegend zwischen strategischen und operativen Wandel differenziert werden, wobei proaktiv, vorausschauend, oder mehr reaktiv, abwartend, agiert werden kann. Hinsichtlich der Stärke bzw. des Umfangs der jeweiligen Veränderungen kann ferner zwischen Wandel geringen Ausmaßes oder tiefgreifenden Wandel differenziert werden.[27] In Anlehnung an Levy und Merry (1986, :9) soll in dieser Arbeit von tiefgreifendem Wandel dann gesprochen werden, wenn einschneidende paradigmatische Änderung des organisatorischen Bedeutungssystems erfolgen, die einen Bruch mit der Vergangenheit darstellen. Tiefgreifender Wandel schließt dabei den Wandel von Konfigurationen unter Einbeziehung von Bestandteilen der Oberflächen- und Tiefenstruktur ein. Wandel geringeren Ausmaßes ist lediglich durch inkrementale Modifikationen ohne Änderung des grundlegenden organisationalen Sinn-/Bedeutungssystems (wie z. B. der Organisationskultur, der strategischen Ausrichtung oder elementarer Prozesse bzw. Strukturen) gekennzeichnet. In Abgrenzung zu Veränderungen, die sich spontan und unvorhersehbar aus verschiedensten Impulsen heraus „ergeben“[28], wird analog zu dem eingeführten Begriffsverständnis von „Strategien“ von „geplantem“ Wandel gesprochen, sofern dieser bewusst initiiert und gesteuert wird. Wohlwissend, dass komplexe Systeme und Prozesse nicht vollkommen plan- und beherrschbar sind, werden Veränderungen hierbei intentional nach zugrundeliegenden Handlungsorientierungen als ex-ante Modelle eines Zielzustandes herbeigeführt.[29] Prozesse organisationalen Wandels sind zunächst sich permanent vollziehende, fortlaufende

[23] Vgl. für eine Übersicht zu verschiedenen Geschäftsentwicklungsstrategien u. a. Herbert und Deresky, 1987.

[24] Sollen nachfolgend Prozesse als Folge von Entscheidungen, Handlungen und Interaktionen bilden, die sich auf die Erfolgspotentiale einer Unternehmung beziehen bzw. organisationale Fähigkeiten betreffen (vgl. Bamberger und Wrona, 2012a:415.).

[25] Vgl. zu Merkmalen und Bedeutung der Zeitdimension in Management- und Organisationsansätzen Perich, 1992:216-303.

[26] Vgl. ausführlich Gomez und Müller-Stewens, 1994 und Kirsch, 1992:133ff..

[27] Vgl. zu relevanten Dimensionen des Ausmaßes von Wandel z. B. Levy und Merry, 1986:9, Staehle, 1999: 900-902, Vahs und Weiand, 2010:3, Kirsch et al., 1979:17ff..

[28] Vgl. zum Begriff des „emergenten“ Wandels Krüger, 2004: 1606.

[29] Die Frage nach dem Maße der intentionalen Planbarkeit von Prozessen des Wandels hinsichtlich zukünftiger Entwicklungen und Zustände führt zu verschiedenen theoretischen Erklärungsansätzen zum Zustandekommen organisationalen Wandels zurück und stellt die klassische Dualität von Volun-

Kennzeichnung von außen nach innen

Perich (1992:133f.)	Krüger (1994:358ff.)	Tushman/Romanelli (1985:179f.)
Funktionale Prozesse (Rationalisierungspotenzial, Strukturen, Systeme)	Strukturen, Prozesse, Systeme, Realisationspotenzial	Steuerungs- und Managementsysteme
Zwecke und Ziele (Strategien und politische Prozesse)	Strategie	Grundlegende Organisationsstruktur (formale Hierarchie etc.)
Soziale Prozesse (explizite Kultur)	Fähigkeiten, Verhalten	Intraorganisationale Machtverteilung
Prädispositionen (Implizite Kultur)	Werte und Überzeugungen	Unternehmensstrategie
Identität		Kernwerte- und Überzeugungen

Abbildung 2: Schichtenmodelle des organisationalen Wandels (Quelle: Adaptiert aus Bamberger und Wrona, 2012a:497).

Prozesse. Allerdings ist es möglich, hieraus kontextabhängig einzelne „Episoden“ der Veränderung, im Sinne eines durch einen Anfang und ein Ende definierten Abschnittes, herauszuheben.

Bei der Betrachtung einzelner Episoden von Entwicklungspfaden ist ferner eine Differenzierung von „evolutionärem“ und „revolutionärem“ Wandel hinsichtlich der Dynamik von Veränderungsprozessen bedeutend. Bezug nehmend auf verschiedene Verlaufsformen[30] kann Wandel danach in der Art sukzessiver, teilweise parallel verlaufende Aktivitäten gradueller Veränderung (inkrementaler, evolutionärer Wandel) eher kontinuierlich vollziehen oder sprunghaft erfolgen (revolutionärer Wandel). Ein traditionelles Modell zur Beschreibung der Unternehmensentwicklung durch ein Alternieren von Phasen des revolutionären und inkrementalen Wandels wurde von Tushman und Romanelli (1985, 1994) mit der Theorie des unterbrochenen Gleichgewichts „Punctuated equilibrium“ vorgestellt. Der Theorie zufolge ist die Entwicklung einer Unternehmung typischerweise durch lang andauernde Phasen der „Konvergenz“, unterbrochen von kurzen, krisenhaften Phasen des „Umsturzes“, gekennzeichnet (vgl. Bamberger und Wrona, 2012a:499f.).

tarismus und Determinismus in den Mittelpunkt. Während voluntaristische Ansätze eine intentionale Gestaltbarkeit von Veränderungsprozessen unterstellen, wird diese in deterministisch orientierten Organisationstheorien als zumindest stark eingeschränkt angesehen, wobei die Entwicklung maßgeblich durch die strukturellen Bedingungen des Kontext bestimmt wird. Auf die Planbarkeit und Möglichkeit zur Steuerung von Wandel ist in Abschnitt 2.2.3 zurückzukommen, wenn Wandelprozesse aus handlungsorientierter Perspektive betrachtet werden (vgl. als Überblick zu verschiedenen Konzeptionen z. B. die Diskussion zur Planbarkeit des Wandels von Bamberger und Wrona, 2012a:507-509; 417).

[30] Es existiert eine Vielzahl an Ansätzen zur Beschreibung der Unternehmensentwicklung bzw. einzelner Objektbereiche und Erfassung entsprechender Entwicklungspfade. Als Überblick zu verschiedenen Modellen siehe etwa Perich, 1992:160ff., Gomez und Müller-Stewens, 1994.

2.2.2. Rückzugsstrategien als Weg zur strategischen Erneuerung

Ungeachtet ihrer starken praktischen Relevanz werden Rückzüge als eigenständiges strategisches Werkzeug zur Implementierung strategischen Wandels bzw. Teil des Wandlungsprogramms[31] zur langfristigen Neuausrichtung des Geschäftsportfolios[32] erst in jüngster Zeit stärker thematisiert.[33]

Entgegen dem oft negativ behafteten Image in der Art einer Niederlage sowie der Auffassung als bloses Spiegelbild zu M&A-Transaktionen und eine Maßnahme der Unternehmens-Restruktuierung[34], sollen mit dieser Arbeit Rückzüge unter dem Gesichtsaspekt strategischer Wandelprozesse zur Anpassung an dynamische Marktverhältnisse beschrieben und unter ihrem Potential zur unternehmerischen „strategischen Erneuerung“[35] untersucht werden.[36]

Damit soll auch dem postitulierten Anspruch von Rückzügen als bewusste Handlungsalternative zur Wertsteigerung, im Sinne des Titel einer Fachzeitschrift für Unternehmens-

[31] Sei hier die Gesamtheit abgestimmter Einzelprojekte bzw. Maßnahmen, u. a. technischer, organisatorischer, personeller und finanzieller Art, hinsichtlich eines größeren Wandlungsvorhabens verstanden (vgl. Krüger, 2002:52).

[32] Als Zusammenfassung aller strategischen Geschäftsfelder einer Unternehmung und ihrer Beziehungen (vgl. Bamberger und Wrona, 2012a:111).

[33] Siehe dazu insb. Brauer, 2006; Decker und Mellewigt, 2007a,b, 2012.

[34] Vgl. hierzu insb. die in der Literatur verbreitete Kategorisierung nach Bowman und Singh (Online-Appendix 1), wonach Rückzüge in Form von „Desinvestitionen“ als ein Element im globalen Feld der Unternehmensrestrukturierung verortet werden.

[35] Unter dem Begriff der „Erneuerung“ sind zunächst allgemeine, bewusst hervorgerufene Veränderung zur Erhöhung des organisationalen Fähigkeitspotentials und Weiterentwicklung der Organisation zur höheren „Reife“, zu verstehen (vgl. Waterman, 1988). Unter einer „Strategischen Erneuerung“ seien hier Wandlungsprogramme und anschließende Veränderungsprozesse subsumiert, die speziell auf die langfristige Veränderung von Erfolgspositionen und/oder Erfolgspotentialen gerichtet und damit von operativen Veränderungen abzugrenzen sind (vgl. Krüger, 2002:51ff.).

[36] Vgl. hierzu auch die begriffliche Unterscheidung im englischen Sprachraum von sog. „strategic business exits“, die einen Strategiewandel implizieren, in Abgrenzung zu „status-quo preserving business exits“ (Decker und Mellewigt, 2012, 2007b).

führung „breaking up is good to do“ (Anslinger et al., 1999), Rechnung getragen werden.

Rückzugsstrategien seien hier als globale Aktivitätsstrukturen zur Verminderung des Engagement einer Unternehmung in produkt-/marktbezogenen Aktivitäten verstanden[37], welche sich auf alle eingeführten Ebenen von Strategienarten (vgl. Abschnitt 2.1.2) beziehen und kontextabhängig verschiedene Formen[38] annehmen können.

Motive zur Verfolgung von Rückzugsstrategien sind zunächst extern bedingt und in der Branchenstruktur[39] bzw. Merkmalen des Marktes und insbesondere dem Verhalten von Wettbewerbern zu sehen. Veränderungen in der Unternehmensumwelt können außerdem Anforderungen an die Ressourcenposition des Unternehmens, z. B. hinsichtlich der verwendeten Technologie, verfügbarer Finanzmittel oder auch Markt-/Länderkenntnisse, stellen, die nicht erfüllt werden können oder eine Erfüllung nicht vorgesehen ist. Wesentliche Motive für einen Produkt-/Marktrückzug können zudem in einer nichtzufriedenstellenden Geschäftsperformance oder Marktposition – auch geknüpft an Portfoliobetrachtungen, sowie Aspekten des Bereichs der Unternehmensführung, wie Wechsel der Führungsspitze oder Einflüsse seitens der Shareholder, z. B. aufgebauter Druck durch Bewertungsabschläge stark diversifizierter Unternehmen, liegen. Nicht zuletzt können strategische Gesichtspunkte eine wichtige Rolle spielen und Rückzugsmotive u. a. in einem Strategiewechsel begründet sein. Während sich Unternehmen reaktiv zurückziehen können, z. B. aufgrund einer sinkenden Geschäftsperformance, können Rückzüge mithin auch Ausdruck vorausschauenden Verhaltens sein und proaktiv vollzogen werden.[40]

In dieser Hinsicht werden in der Literatur zu Rückzugsprozessen unterschiedliche Transformationstypen[41] mit strategischer Implikation thematisiert. Basierend auf den Arbeiten von Byerly et al. (2003), soll eine Typisierung von Rückzugsstrategien zur „Refokussierung“ bzw. „Repositionierung“ von Geschäftsaktivitäten einer Unternehmung erfolgen. Unter einer Refokussierung wird die Aufgabe vorwiegend peripherer Aktivitäten zur Stärkung des Kerngeschäfts[42], oft in Verbindung mit einer Änderung der Strategie der Spezialisierung und Diversifikation[43], verstanden, wohingegen eine Repositionierung auf das Einschlagen eines „neuen Weges“ im Sinne eines strategischen Richtungswandel bzw. der Etablierung eines neuen Kerngeschäfts abzielt. Die beiden Typen sind grundsätzlich in ihrem Zusammenwirken zu betrachten, eine strategische Neuausrichtung des Kerngeschäfts kann mit der Reduktion einzelner Aktivitäten verbunden sein (vgl. z. B. Markides, 1993; Johnson, 1996; Byerly et al., 2003; Hoskisson et al., 2005; Decker und Mellewigt, 2007b, 2012).

Aufbauend auf den Ergebnissen empirischer Forschungsarbeiten betonen Brown und Eisenhardt (1998, 1999) die Bedeutung von Rückzugsstrategien zum Bestehen im Wettbewerb – besonders in hochdynamischen Märkten – mit dem Prozessansatz des „Patching“[44], als organisationale und strategische Routinen bzw. „dynamic capabilities“[45] zur kontinuierlichen und proaktiven Anpassung an Turbulenzen der Unternehmensumwelt.

Dabei legen sie eine integrative Perspektive von Strategie- und Wandelprozessen zugrunde, um die „Dynamik“ des Strategieprozesses besser abzubilden, und kritisieren traditionelle Annahmen einer Gleichgewichtsvorstellung[46] zur Betrachtung von Verlaufsformen organisationalen Wandels: „Strategy is the result of a firm’s organizing to change constantly and letting a amicoherent strategic direction emerge from that organization.” (Brown und Eisenhardt, 1998:7).

[37] Rückzugsstrategien stehen dabei in engem Bezug zu den in der Wissenschaft sowie Praxis vielfach verwendeten Begriffen der Desinvestitionsstrategien, Schrumpfungsstrategien oder auch Marktaustrittsstrategien, sind jedoch nicht zwangsläufig synonym zu verwenden. In Anlehnung an Überlegungen zu verschiedenen Rückzugsdimensionen, können Rückzüge neben der Aufgabe von Produkt-/Marktaktivitäten auch die Rücknahme von Marktpositionen einschließen, ohne dass a priori eine vollständige Aufgabe unterstellt bzw. eingeschlossen ist, oder sich auf der Ebene einzelner Wertaktivitäten vollziehen (vgl. Bamberger und Delic, 2010). Rückzugsstrategien werden nicht nur unter Bedingungen schrumpfender Märkte gewählt und werden in diesen nicht zwangsläufig verfolgt (siehe für einen umfassenden Literaturüberblick der Managementforschung zu schrumpfenden Märkten u. a. Koponen und Heikki, 2009, als Überblick zu Schrumpfungsstrategien z. B. die Ausführungen von Welge und Al-Laham, 2008:619-632. und dabei insb. die Studien von Harrigan, 1980, 1981 und Harrigan und Porter, 1983).

[38] Zur Unterscheidung verschiedener Formen von Rückzugsstrategien siehe z. B. Bamberger und Delic, 2010; Harrigan, 1982; Mitchell, 1994 und speziell zu Formen des Sell-off, Spin-off, Carve-out und Buy-out auch Stienemann, 2003.

[39] Branchen sind hierbei der Definition von Porter (2008) folgend als „eine Gruppe von Unternehmungen, die Pro-dukte herstellen, die sich gegenseitig nahezu ersetzen können“ (Porter, 2008:37) zu verstehen.

[40] Vgl. Bamberger und Wrona, 2012a:212-218, Johnson, 1996; Decker und Mellewigt, 2012, 2007a; Markides, 1993; Brauer, 2006 sowie speziell zu unterschiedlichen Systematisierungen von Desinvestitionsmotiven Dohm, 1988 und Ossadnik, 2002.

[41] Transformation als Kennzeichnung eines umfassenden, tiefgreifenden Wandels (vgl. u. a. Levy und Merry, 1986).

[42] Als „Kerngeschäft“ einer Unternehmung sind zentrale bzw. ursprüngliche Tätigkeitsfelder zu sehen, welche meist hauptsächlich zum Erlös beitragen und die Rendite sichern (vgl. z. B. die Definition von McGranth, 1995).

[43] Siehe zur Änderung der Diversifikationsstrategie im Rahmen der Unternehmensfokussierung u. a. Hoskisson und Johnson, 1992; Hoskisson et al., 1994 und Byerly et al., 2003.

[44] „Patching“ wird als strategisches Vorgehen verstanden, mit dem Geschäftseinheiten permanent neu formiert und fokussiert werden können – je nach. Patching-Prozesse können unterschiedliche Formen der Erweiterung, Spaltung, Transferierung, Aufgabe oder Kombinierung einzelner Geschäftsaktivitäten annehmen: “Patching is the strategic process for routinely remapping businesses to take advantage of changing market opportunities. It can take the form of adding, splitting, transferring, exiting, or combining chunks of businesses.” (Brown und Eisenhardt, 1999:74). Martin und Eisenhardt (2004) entwickeln weiterführend das Konzept einer „recombinative organizational form“ durch Fokussierung verschiedener „corporate dynamic capabilities“ (wie u. a. Patching-Prozesse) und einer „modular organizational structure“ als Strategie in schrumpfenden Märkten zur Erschließung neuer Marktchancen: “Corporate entrepreneurship emphasizes exiting from declining markets, while simultaneously capturing new opportunities by recombining resources deployed in these declining markets in novel ways.“ (Martin und Eisenhardt, 2004:176).

[45] Vgl. Teece et al., 1997; Eisenhardt und Martin, 2000; Winter, 2003.

[46] Vgl. hierzu die Perspektive von Miller und Friesen, 1980 und Tushman und Romanelli, 1985, 1994 zum Verlaufsmodell des „Punctuated equilibrium model of organizational transformation“.

2.2.3. Prozesse des Wandels aus handlungsorientierter Perspektive

Die Analyse der strategischen Prozesse, in denen Veränderungen hervorgebracht werden, führt zur handlungsbezogenen Prozessperspektive. Dabei ist von besonderem Interesse, inwieweit Veränderungsprozesse intentional herbeigeführt und gesteuert werden können und welche Einflussfaktoren für den Entwicklungsverlauf prägend sind.

Zur Klärung des spezifischen Zustandekommens komplexer Wandel-Phänomene und dem Maße der Steuerbarkeit der Unternehmensentwicklung soll in dieser Arbeit a-priori weder eine rein voluntaristisch orientierte Sichtweise, wie v. a. in planungsorientierten (normative) Konzeptionen vorherrschend, noch deterministische Perspektive eingenommen werden, die sich insbesondere in situativen Ansätzen (Kontingenztheorien), sowie evolutionstheoretischen Überlegungen, speziell dem Population Ecology-Ansatz wiederspiegelt. Den folgenden Überlegungen liegt eine intermediäre Position zugrunde, wobei Organisationen zwar der Evolution unterliegen, die Entwicklung aber durch intendierte Wahlakte gemäß eigener Ziele prinzipiell veränderbar ist.

Allerdings wird die Fähigkeit zu einer intentionalen Steuerbarkeit des Wandels als begrenzt betrachtet (vgl. hierzu die Sichtweise des sog. „gemäßigten Voluntarismus“, Kirsch et al., 1979; Kirsch, 1996). Diesbezüglich spielen Ursachen dafür, dass Organisationen trotz Anpassungserfordernissen, nicht in der Lage sind, sich (grundlegend) zu wandeln, eine maßgebliche Rolle, worauf im Folgenden näher eingegangen werden soll.

Organisationale Trägheit und strategische Rigidität

Organisationale Veränderungsprozesse von Systemen lassen sich in einem Spannungsfeld von Triebkräften des Wandels und Gegenkräften der Beharrung betrachteten.[47] Als Merkmal, notwendig erachtete Änderungen von Systemzuständen nicht vorzunehmen, hat sich der ursprünglich aus der Physik stammende Begriff der „Trägheit“[48] in der Strategie- und Organisationsforschung etabliert. Organisationale Trägheit impliziert ein Festhalten an konstituierenden (strategischen) Mustern und kann in verschiedenen Ausprägungsformen auf den Wandlungsprozess einwirken.

Trägheit kann sich darin äußern, dass eine Selbstreflexion zum Erkennen eines Wandlungsbedarfs und/oder die anschließende Umsetzung von Veränderungsmaßnahmen verhindert bzw. verzögert wird oder in gleicher Weise beeinflussen, dass Pläne zur Veränderungen zwar autorisiert, allerdings nicht wie vorgesehen implementiert werden bzw. auf Widerstände von Betroffenen stoßen und nicht zu realen Verhaltensänderungen führen.

Für die weiteren Überlegungen sei an dieser Stelle betont, dass Trägheit in dieser Arbeit vornehmlich nicht negativ assoziiert und klar von dem in der Literatur oft synonym verwendeten Widerständen abzugrenzen ist. Unter Widerstand sind mentale Barrieren zu verstehen, die sich in einer aktiven oder passiven Ablehnung von Veränderung und bewusst initiierter Prozess zur Aufrechterhaltung des „Status quo“ zeigen. Widerstände können auf der Ebene einzelner Personen als auch auf Gruppen- und Unternehmensebene auftreten und sind häufig politischer, rationaler oder emotionaler Natur.[49] Trägheit hingegen bezeichnet einen unbewussten Prozess, der in einer mangelnden Fähigkeit zum Wandel begründet liegt (vgl. hierzu das „3-W“-Modell in Abschnitt 2.2.3.).

Speziell auch unter dem Blickwinkel der Gestaltung des Wandels von Systemen und spezifischer Problembereiche knüpfen Ansätze zur Erklärung von Widerstand und Trägheit an unterschiedlichen Konzeptionen der Organisationstheorie an. Einen klassischen Ausgangspunkt zur Untersuchung von Trägheitsphänomenen bilden Aussagen des Population Ecology-Ansatzes. Diese werden nachfolgend vorgestellt und durch Kernargumente zusätzlich ausgewählter Konzepte und Theorien zur Thematik ergänzt.

Ausgehend von der Fragestellung nach der Diversität von Organisationsformen thematisieren Hannan und Freeman (1984, 1989) mit dem Population Ecology-Ansatz Phänomene „organisationaler Trägheit“ im Theorierahmen von zentralen Erklärungsmechanismen der Evolutionsbiologie für die Veränderungen innerhalb Organisationspopulationen.[50]

Trägheitskräfte werden innerhalb der Selektions-Adaptions-Debatte als wesentliche Begründung für die unterstellte begrenzte Anpassungsfähigkeit von Organisationen an eine feindliche Umwelt und Voraussetzung einer negativen Selektion im Konkurrenzkampf nach dem Mechanismus des „survival of the fittest“ angesehen. Gleichzeitig kann Trägheit aber auch existenzfördernd sein und danach als Ergebnis von Selektionsprozessen und zentrales Argument zur Erklärung der Stabilität von Populationen ausgelegt werden.[51] In Bezug auf Aspekte des organisatorischen Wandels wird Trägheit ebenso für die Sprünge in der Organisationsentwicklung im Rahmen

[47] Vgl. hierzu grundlegend die gruppendynamische Theorie des sozialen Kräftefeldes von Lewin. und für einen umfassenden Überblick möglicher Erklärungsfaktoren organisatorischen Wandels z. B. Perich, 1992: 159.

[48] Vgl. u. a. Huff et al., 1992: „Absent to other forces, inertia describes the tendency to remain with the status quo and the resistence to strategic renewal outside the frame of current strategy. “. Im deutschen Sprachgebrauch finden sich auch Begriffe wie „Rigidität“, „Immobilität“, „Verkrustung“ oder „Konservatismus“.

[49] Vgl. u. a. Schäcke, 2006:157, Niemeier, 2000:251ff.; Schreyögg, 1998:489ff., Doppler und Lauterburg, 2008:336; Kieser et al., 1998:123f., speziell zur Unterscheidung von Widerstandsarten siehe z. B. Vahs, 2007:20, Eschenbach, 1995:91.

[50] Evolutionstheoretische Ansätze lehnen sich an Darwins synthetische Evolutionsbiologie an und legen dabei ein deterministisches Verhältnis von Organisationen und Umwelt zugrunde (vgl. Darwin, 1859, für eine Übersicht der synthetischen Evolutionsbiologie siehe Kieser und Woywode, 2006:309f.). Analyseebene der klassischen Forschung ist die Population von Organisationen, die durch einen gemeinsamen „Blueprint“ definiert wird (vgl. Hannan und Freeman, 1977:936). Die Veränderung innerhalb Organisationspopulationen basieren demnach auf den Evolutionsmechanismen der Variation, Selektion und Retention.

[51] Die Begründung hierfür fußt auf der Annahme, dass die Umwelt bzw. Ressourcengeber von Organisationen Reproduzierbarkeit ihrer Leistung fordert. Diese resultiert aus der Gewährleistung eines konstanten Qualitätsniveaus („Reliability“) sowie der Fähigkeit zur Rekonstruktion der erstellten Leistung („Accountability“). Dazu werden Systeme, im Kern standardisierte, häufig institutionalisierte Routinen, entwickelt, die Stabilität von Strukturen und Verhaltensweisen schaffen (vgl. Hannan und Freeman, 1984).

der erwähnten Konzeptualisierung des „Punctuated equilibrium“ verantwortlich gemacht.

Darüber hinaus unterscheiden *Hannan/Freeman* zwischen absoluter, auf die Veränderungsrate im Sinne der Veränderungshäufigkeit von Organisationen bezogener, und relativer Trägheit. Relative Trägheit beschreibt die Veränderungsgeschwindigkeit von Organisationen im Verhältnis zur Veränderungsgeschwindigkeit der umgebenden Umwelt. Diese soll in dieser Arbeit auch im Mittelpunkt stehen, da davon ausgegangen wird, dass Organisationen grundsätzlich zur Anpassung fähig sind.

Organisationale Trägheit wird durch eine Reihe organisationinterner und -externe Prozesse hervorgerufen (siehe auch Abbildung 3). Zu internen Faktoren zählen „Sunk Costs“[52], Unvollständige Informationen des Managements infolge fokussierter Informationssysteme, das interne politische Machtgefüge sowie das historisch gewachsene Wertesystem. Als externe Ursachen werden Markteintritts- und Austrittsbarrieren sowie ähnlich der internen Informationsrestriktion der Zugang zu Informationen, die erworbene Legitimität bzw. Reputation der Umwelt und schließlich die Unvollkommenheit kollektiver Entscheidungsprozesse gesehen.

Nach einer Hypothese im Zusammenhang mit diesem Wirkungsgefüge steigen Trägheitskräfte mit der Lebenszyklusphase (Alter), der Größe und Komplexität der Organisation an.[53]

Neben dem im Population Ecology-Ansatz integrierten und empirisch fundierten Trägheitskonzept beschäftigen sich vor allem der organisatorische Konservatismus und Forschungen hinsichtlich einer zeitlichen Dimension auf historische Prozesse, wie insbesondere die Pfadtheorie und Ansätze zu Kernkompetenzen mit Trägheitsideen.

Aus dem Blickwinkel kognitiv-interpretativer Theorieströmungen[54] bilden Merkmale von Wissensstrukturen[55] einzelner Organisationsmitglieder sowie auf kollektiver Ebene der Gruppe bzw. Organisation im Sinne kognitiver Verfestigungen und Phänomene selektiver Wahrnehmung Ursachen von Trägheit. Kieser et al. (1998) verwenden den Begriff des „organisatorischen Konservatismus“ als Bedürfnis nach Kontinuität und Sicherheit, das sich auf Strukturen, Abläufe, Strategien und Verhaltensweisen von Organisationsmitgliedern beziehen kann. Unternehmensinterne und -externe Ursachen von Konservatismus sind dafür verantwortlich, dass Veränderungen in Organisationen nicht vollständig oder zeitnah umgesetzt werden. Unternehmensinterne Ursachen werden weiterhin in Widerstände und Trägheit auf Verhaltensebene, im Sinne eines fachlichen oder kognitiven Unvermögens einzelner Personen zur Veränderung, wie Systemebene, in Anlehnung an die Überlegungen des dargestellten Population Ecology Ansatzes, unterschieden.[56]

Hilfreich bei der Betrachtung von Bedeutung und Einfluss der Historizität bzw. frühere Entscheidungen und Ereignisse, bezugnehmend auf interorganisationale Ursachen des Trägheitsphänomens, ist das Konzept der Pfadabhängigkeit und der damit verbundene Begriff des „Lock-In-Effekts“ - einen Zustand, in dem sich unter bestimmten Bedingungen ein ineffizientes Gleichgewicht einstellt, welches unwahrscheinlich wieder verlassen wird. Ursprünglich hat die ökonomisch motivierte Pfadtheorieforschung untersucht, warum sich inferiore Technologien durchsetzen[57], inzwischen hat sie zur Erklärung vielfältiger Phänomene, u. a. eine Ausdehnung auf Verhaltensweisen und Strategien, an Bedeutung gewonnen. Das Konzept betont die Historizität von Institutionen, wobei angenommen wird, dass durch vergangene Entscheidungen oder Ereignisse („Critical Juncture“), die durch Mechanismen „positiver Rückkopplung“[58] gesteuert bzw. rekursiv stabilisiert werden, ein „Momentum“ generiert wird, das den (strategischen) Handlungsspielraum nachhaltig limitiert und pfadbedingte Rigidität bewirkt, die sich in der trägen Gleichförmigkeit des Entscheidungsverlaufs äußert.[59]

[52] Entstehen durch die Mobilisierung von Ressourcen in Form irreversibler Investitionen in Anlagen, Ausstattung, Zeit oder persönlichen Einsatz (vgl. Hannan und Freeman, 1977:931).

[53] In einer Erweiterung ihres Grundkonzepts zur Erklärung des Trägheitsphänomens haben Hannan/Freeman die wesentlichen Einflussfaktoren auf Trägheit und verbundene Wirkungszusammenhänge mit dem Fokus auf Wandel und Scheitern von Organisationen konzeptualisiert. Diese Faktoren und Beziehungen wurden von anderen Forschern sowohl theoretisch, als vorwiegend auch empirisch aufgegriffen, untersucht und erweitert (vgl. als Übersicht ausgewählter empirischer Studien zu den Einflussfaktoren Kelly und Amburgey, 1991:593).

[54] Vgl. zur Unterscheidung verschiedener Theorieansätze überblicksartig z. B. Bamberger und Wrona, 2012a:66ff.. Kognititv-interpretative Ansätze sollen danach im weiteren Sinne verstanden werden.

[55] Der Begriff „Wissensstruktur“ soll eine Menge geordneter und auf unterschiedliche Weise miteinander verknüpfter Kognitionen bezeichnen, die das Verhalten von Menschen speziell dadurch beeinflussen, dass sie die Aufmerksamkeit und Wahrnehmung und damit Erfassung von Problemen sowie Suche nach Lösungen lenken, Informationen Bedeutung geben und mithin als interpretative Schemata wirken. Bestandteile von Wissensstrukturen können dabei sowohl (deskriptive) Kognitionen, mit denen Informationen über die Realität repräsentiert werden, als auch normative bzw. präskriptive Merkmale, wie Werte oder Attitüden bilden. Wissensstrukturen sind das Ergebnis permanenter Wahrnehmungs-, Interpretations- und Lernprozesse. Sie repräsentieren dabei auch verschiedene Lebenswelten des Individuums und sind als mögliche Bezugspunkte für Wertesysteme, Überzeugungen und Normen zu sehen. Im Rahmen des strategischen Managements beziehen sich u. a. „Kognitive Landkarten“ (Axelrod, 1976), „Paradigmen“ (Halme, 2002), „dominante Logiken“ (Prahalad und Bettis, 1986, Bettis und Prahalad, 1995) oder „interpretative Schemata“ (Greenwood und Hinings, 1988) auf wichtige Komponenten von Wissensstrukturen.

[56] Vgl. zum organisatorischen Konservatismus vornehmlich Kieser et al., 1998:121 ff., für Details zu Verhaltens- und Systemwiderständen weiterhin Kieser et al., 1998: 124 f. und 127.

[57] Vgl. insb. David, 1985: 332 ff. und Arthur, 1989.

[58] Als selbstverstärkende Effekte wie z. B. infolge Skalenerträge, Lerneffekte oder Koordinations- und Komplementaritätseffekte (vgl. David, 1985; Arthur, 1989; Schäcke, 2006:36 ff.; Schreyögg et al., 2003: 262 f.; Ackermann, 2001: 16 ff. und Ackermann, 2001 S. 59 ff.). Arrow, 2003 erweitert das bestehende Pfadabhängigkeitskonzept um „Constant Returns“, für ihn ist ausschließlich die Quasi-Irreversibilität von Investitionen der kritische Faktor einer Pfadabhängigkeit. Mahoney, 2000 benennt utilitaristische, funktionale, machtbasierte und legitimatorische Gründe als Ursachen für institutionelle Reproduktion.

[59] Vgl. zur Pfadabhängigkeit allgemein Schreyögg und Sydow, 2003; Sydow et al., 2005: 5 ff., Schreyögg et al., 2003: 259 ff., Beyer, 2006; Ackermann, 2001: 9 ff.; zu den Eigenschaften pfadabhängiger Prozesse insb. Schäcke, 2006: 29 f.; Ackermann, 2001: 19 ff. sowie zur Entwicklung von Pfaden insb. Sydow et al., 2005: 8 ff., Schreyögg et al., 2003 267ff..

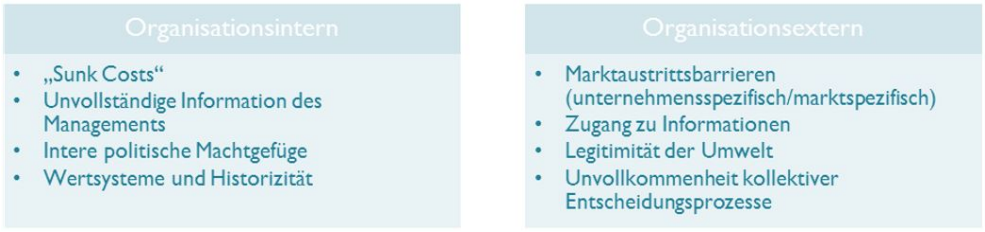

Abbildung 3: Ursachen organisationaler Trägheit (Quelle: in Anlehnung an Hannan und Freeman, 1977:931f., Hannan und Freeman, 1984:149 und Hannan und Freeman, 1989:67ff.).

Basierend auf der Tradition des „Ressource-Based-View“ werden speziell auch „Kernrigiditäten“ durch historisch gewachsene Lern- und Verfestigungstrajektorien erklärt, welche Organisationen daran hindern sich den geänderten Marktverhältnissen anzupassen.[60]

Daneben sind auch andere klassische Sichtweisen der Ursachen von Trägheit, wie die Industrieökonomik durch Betonung von Markteintritts- und austrittsbarrieren in Bezug auf extern-orientierte Kontextfaktoren von Branchenmerkmalen oder explizite Betrachtung von Wandelprozessen als politische Prozesse hinsichtlich der (pluralistischen) Struktur und Macht von Führungssystemen, möglich.

Auf das einleitend angesprochene Kräftefeld von Trägheitskräften und Treibern des Wandels zurückkommend, sei an dieser Stelle auf die Ambivalenz dargestellter Trägheitsfaktoren hingewiesen. So können mikropolitische Taktiken auch gezielt zur Promotion von Wandel eingesetzt, der institutionelle Kontext kann nicht nur verfestigend wirken, sondern auch den Rahmen zur Diffusion von Innovation (Pfeffer und Salancik, 1978) bzw. die zur grundlegenden Veränderung nötige Ressourcenbasis bilden (Powell, 1991). Die unter dem Gesichtsaspekt von Trägheit behandelten Faktoren besitzen insofern einen dualen Charakter, indem sie Wandel auf der einen Seite begrenzen, ihn andererseits aber auch begünstigen oder gar erst ermöglichen.

Management des Wandels

Prozesse des Wandels bedürfen eines Managements, weil sie sehr komplexe Prozesse darstellen, die eine Führung für einen erfolgreichen und effizienten Verlauf erfordern.

Ausgehend von der klassischen Organisationsentwicklung (vgl. Lewin, 1951; Blake und Mouton, 1964; Peters und Waterman, 1982) werden in der BWL Vorschläge zu strategisch ausgerichteten Konzepten und Umsetzungsimpulsen gemacht, die geplanten Wandel von Organisationen erzeugen und gestalten können. Seit Mitte der 80er Jahre werden psychologische, soziologische und ökonomische Komponenten des Wandels in der Literatur unter dem Stichwort „Change Managements“[61] gebündelt.

Ein solches „Change Management“ bezieht sich allgemein auf die zielgerichtete Analyse, Planung, Realisierung und Evaluierung von Wandlungsprozessen in Organisationen (vgl. Bamberger und Wrona, 2012a:513, Müller-Stewens und Lecher, 2001:392-404), wobei schwerpunktartig spezifische Episoden in der Phase der Implementierung von tiefgreifenden Veränderungen betrachtet und speziell die Handhabung von Problembereichen der „politisch-verhaltensorientierten“ und „wertmäßig-kulturellen“ Dimension im Mittelpunkt stehen.

Nach Krüger, 2002:68ff.) bilden Wandlungsbedarf, Wandlungsfähigkeit und Wandlungsbereitschaft Bestandteil eines Analyserasters der geplanten strategischen Erneuerung. Mit dem „3-W“-Modell stellt er hiezu ein generisches Rahmenkonzept vor, das diese drei Problemfelder als gedankliche Koordinaten zueinander in Beziehung setzt (vgl. Abbildung 4). Ausgangspunkt von Wandelprozessen bildet zunächst der Wandlungsbedarf als Ausmaß der sachlich notwendigen Veränderung, die intern oder extern getrieben sein kann und bestenfalls proaktiv angegangen wird. Die Wandlungsbereitschaft beschreibt die Einstellungs- sowie Verhaltensakzeptanz der Prozessbeteiligten bzw. -Betroffenen sowie Organisationseinheiten gegenüber den Zielen und Maßnahmen des Wandels (Promotoren vs. Opponenten). Die Wandlungsfähigkeit ist schließlich die auf entsprechenden personen- und sachbezogenen Einflussgrößen basierende Möglichkeit, Prozesse des Wandels erfolgreich und effizient durchführen zu können.

Wandel wird als Daueraufgabe betrachtet, wobei das Change Management insbesondere auf die Beeinflussung von Wandlungsbereitschaft und Wandlungsfähigkeit und damit dem Umgang mit Trägheit und Widerstand abzielt, um Veränderungsprozesse gezielt zu steuern und die Kongruenz zwischen den drei Koordinaten zu erreichen. Den Kernaktivitäten des Change Managements liegen dabei insbesondere soziale Prozesse zugrunde, weswegen Konzeptionen entsprechend durch eine starke individual- und sozialpsychologische Orientierung und den Rückgriff auf andere verhaltenswissenschaftliche Disziplinen geprägt sind (vgl. Krüger, 2002; Doppler und Lauterburg, 2008; Kieser et al., 1998).

[60] Vgl. zum Ressourcenansatz u. a. Barney, 1991; Wernerfelt, 1984, 1995; Petraf, 1993; zum Kernkompetenzansatz insb. Pralahad und Hamel, 1990; Leonard-Barton, 2007; Dougherty, 1995; siehe hierzu auch den Begriff der „competency trap“ (Levitt und March, 1988).

[61] Zur Einordnung und Abgrenzung zum ganzheitlichen Ansatz der „Organisationsentwicklung“ siehe z. B. Doppler und Lauterburg, 2008:89-100 und Vahs und Weiand, 2010:7.

Abbildung 4: Koordinaten geplanten Wandels (Quelle: Krüger, 2002).

2.2.4. Bezugsrahmen strategischer Prozesse des Wandels

Organisationaler oder strategischer Wandel ergibt sich als Ergebnis entsprechender strategischer (Entscheidungs-) Prozesse, welche sowohl als kontinuierliche Vorgänge wie auch in Form herauszuhebender Episoden der Veränderung betrachtet werden können, und sind dabei in einem sich selbst permanent wandelnden Kontext verankert.

Um der Vielschichtigkeit strategischer Wandelprozesse und Mannigfaltigkeit möglicher Verlaufsformen gerecht zu werden, sollen dargestellte Perspektiven innerhalb einer offen gestalteten Konzeptualisierung zusammengefügt werden. Dazu wird das allgemeine Rahmenkonzept strategischer Prozesse nach Bamberger und Bamberger und Wrona (2012a, 440ff.) speziell auf das Zustandekommen strategischer Wandelprozesse übertragen. Diese in Abbildung 5 wiedergegebene Modifikation von Merkmalen strategischer Prozesse des Wandels soll als heuristischer Bezugsrahmen der empirischen Analyse zugrunde gelegt und untersuchungsspezifisch ausgefüllt werden.

Prozesse strategischen Wandels lassen sich danach zunächst inhaltlich, anhand angesprochener Objektbereiche sowie Arten des Wandels, klassifizieren und dabei durch allgemeine Prozessmerkmale, wie eine Menge vielfältiger Aktivitäten und ihrer Wirkzusammenhänge, verschiedenen Akteure, Methoden und den Kontext, der sie beeinflusst und auf den sie gleichsam zurückwirken, beschreiben. Merkmale des internen Kontextes, wie konkret die Organisationsstruktur, verschiedene Managementsysteme und die Unternehmenskultur, sowie des externen strukturellen Rahmens (Mikro- und Makroumwelt) und ihre Veränderung können, im Sinne eines wahrgenommenen, sachlich notwendigen Anpassungsverhaltens, Wandlungsprozesse auslösen, aufgrund der angesprochenen Ambivalenz einzelner Faktoren jedoch auch als Ursache von Kräften gegen Veränderung erscheinen.

Verlaufsformen beziehen sich auf ein mögliches ungesteuertes Herausbilden von Veränderungsprozessen oder kennzeichnen insbesondere auch das Zustandekommen eines geplanten Wandels als Gegenstand und Ergebnis intentional gesteuerter Entscheidungs- bzw. Planungsprozesse. Im Rahmen des permanenten „Ongoing Process“[62] kann sich Wandel kontextabhängig und relativ in Form bedeutsamer Episoden revolutionär und/oder inkremental vollziehen.

Unabhängig von Gegenstandsbereichen und der Verlaufsform, lassen sich strategische Wandelprozesse aus verschiedenen Dimensionen untersuchen. Neben der Entscheidungsprozess-Perspektive als zentrale Analysedimension der strategischen Prozessforschung und der Betrachtung strategischer Sachverhalte als Ergebnis von Entscheidungs- bzw. Planungsprozessen (als spezifische[63] Entscheidungsprozesse) seien für Wandelprozesse insbesondere die politisch-verhaltensorientierte sowie wertmäßig-kulturelle Dimension hervorgehoben, womit Aspekte der Tiefenstruktur angesprochen sind. Diese Dimensionen sind speziell auch aus dem Blickwinkel der Gestaltung und Steuerung von Prozessen eines geplanten Wandels im Sinne eines „Change Managements“ relevant, da sie spezifische Problemfelder der Realisierung bzw. Implementierung von Strategien beinhalten, die im Mittelpunkt der Fallstudie stehen.[64]

[62] Als fortlaufender Fluss von Entscheidungen, Handlungen und Interaktionen des organisatorischen Geschehens (vgl. Kirsch, 1997b).

[63] Insofern spezifisch, als dass sie sich durch antizipatives, vorausschauendes und systematisches Vorgehen zur Problemlösung von anderen Arten der Entscheidungsfindung abgrenzen.

[64] Unter dem Gesichtsaspekt, dass Verhalten auf einer informellen Grundlage basiert, wird ferner der Bezug zur Lernperspektive hergestellt, welche in der Betrachtung allerdings nicht vordergründig betrachtet werden soll.

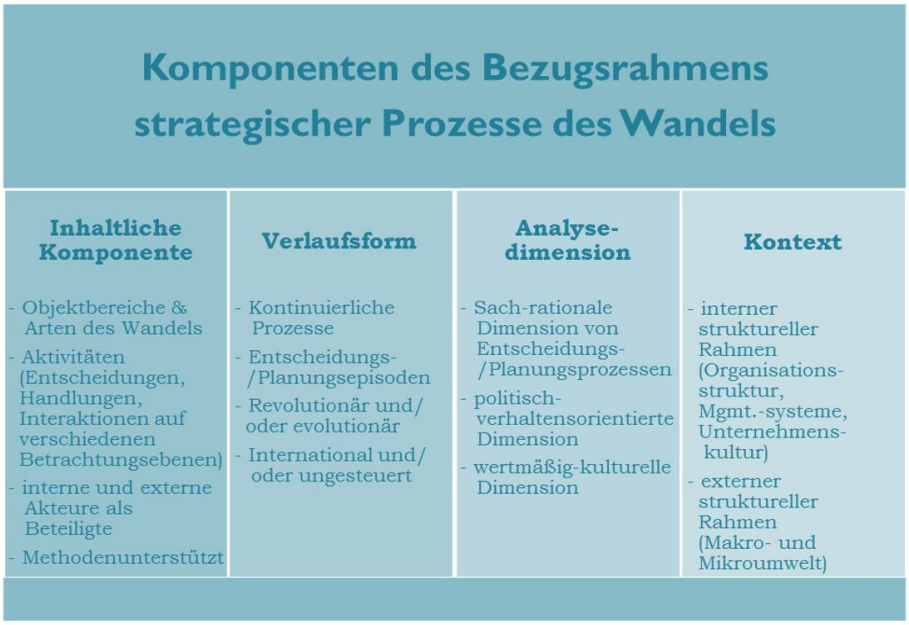

Abbildung 5: Heuristischer Bezugsrahmen zur Analyse von strategischen Prozessen des Wandels (Quelle: Modifiziert aus Bamberger und Wrona, 2012a:441).

3. Fallstudie zur ChemCo

In diesem Kapitel werden die erarbeiteten Grundlagen von Unternehmen im Spannungsfeld von Trägheit und Veränderung empirisch untersucht.

Wie im einleitenden Kapitel 1 ausführlich erläutert, steht die empirische Analyse dabei unter der zentralen Forschungsfrage, ob und warum Unternehmen dazu neigen, beständig gegen Wandel zu sein, und sucht konkret nach den Ursachen für strategische Beharrungstendenzen.

Als Forschungsstrategie wird der qualitative Fallstudienansatz gewählt und hierzu in Kapitel 3.1 zunächst das Forschungsdesign vorgestellt und die Methodik des Forschungsprozesses dargelegt.

Im Anschluss daran erfolgt die Darstellung der Fallstudienarbeit, gegliedert in zwei zentrale Abschnitte:

Zu einem allgemeinen Gesamtverständnis des Forschungsgegenstands – der strategischen Wandlungsprozesse des Halbzeuge-Geschäfts mit PMMA – werden in Kapitel 3.2 zuerst relevante Hintergründe zur Entstehung und Entwicklung dieses Geschäfts der PlastiCo am Anfang bis heute als Bestandteil der operativen Aktivitäten im Großkonzern der ChemCo gegeben. Die Erläuterungen werden ergänzt durch grundsätzliche Informationen zu PMMA-Produkt-/ Marktaktivitäten und produktionstechnische Hinweise, welche beeinflussende interne Kontextfaktoren für den untersuchten Sachverhalt bilden. Ausgehend von verschiedenen Anwendungsbereichen allgemein wird dann speziell auf das Halbzeuge-Geschäft mit Acrylglas im Sanitär-Bereich eingegangen, das im Mittelpunkt der Untersuchung steht.

Daran schließt sich im Schwerpunktkapitel 3.3 die umfangreiche Analyse zur Beantwortung der forschungsleitenden Fragestellung an.

Hierzu werden relevante Wandlungsprozesse von 1986 bis 2008 anhand spezifischer Komponenten beschrieben und auf Ausprägungen und Ursachen von Trägheit geprüft. Diese Analyse gliedert sich wiederum in zwei grundlegende Schritte: Schritt 1 dient zunächst einer allgemein inhaltlichen Kennzeichnung des Untersuchungsgegenstandes (Abschnitt 3.3.1) und beschreibt daraufhin den historischen Verlauf der Geschäftsentwicklung zur Offenlegung potentieller Beharrungstendenzen (Abschnitt 3.3.2). Erhärtet sich der Trägheitsverdacht, schließt in Schritt 2 eine Diagnose an (Abschnitt 3.3.3). Die Ergebnisse der empirischen Befunde werden abschließend zusammengeführt und der Erklärungsbeitrag in Abschnitt 2.2.3 vorgestellter theoretischer Konzeptionen diskutiert (Abschnitt 3.3.4).

3.1. Der Fallstudienansatz als Forschungsstrategie

3.1.1. Eine holistische Einzelfallstudie am Beispiel der ChemCo

Ausgehend von dem zu erforschenden Phänomen wurde das Fallstudiendesign auf eine holistische Einzelfallstudie[65] eingegrenzt und der strategische Wandel des Halbzeuge-Geschäfts mit PMMA der damaligen PlastiCo, heute unter dem Dach der ChemCo, als „kritischer Fall“[66] angelegt.

[65] In Anlehnung an die Fallstudien-Typologisierung nach Yin 2009 stehen prinzipiell vier verschiedene Fallstudientypen zur Verfügung. Der danach für die vorliegende Arbeit relevante Typus einer holistischen Einzelfallstudie untersucht ein einzelnes Analyseobjekt mit angrenzendem Kontext innerhalb eines Einzelfalls (vgl. Yin, 2009:46f.).

[66] Innerhalb der Konstituierung einer Einzelfallstudie erfüllt die als „kritischer Fall“ angelegte Fallstudie die Funktion, Annahmen und Rahmenbedingungen bestehender Theorien zu testen und ggf. neue Perspektiven einzubringen (vgl. Yin, 2009:47, Locke, 2001:97).

In diesem Sinne bildet der Rückzugsprozess aus dem Sanitärgeschäft im Kontext der strategischen Neuausrichtung des Halbzeuge-Geschäfts den Untersuchungsgegenstand, der innerhalb einer retrospektiven Längsschnittstudie[67], basierend auf einer Reihe von problemzentrierten Interviews[68] sowie vielfältigen Dokumentationen und Archivdaten unterschiedlicher Bestände, explorativ untersucht wird.

Auf Grundlage des interpretativen Paradigma wird der historische Entwicklungsverlauf von 1986 bis 2008 unter Zuhilfenahme verschiedener Techniken der Grounded Theory[69] analysiert.

3.1.2. Datenerhebung und- auswertung: Multiple Datenquellen und Triangulation

In Abhängigkeit des Vorwissens und vorhergehenden theoretischen Fundierung werden in der Literatur unterschiedliche, zumeist zirkuläre Vorgehensweisen diskutiert. Nach Yin ist es zulässig eine vorläufige Auffassung über theoretische Konzepte und den Untersuchungszusammenhang zu haben, um diese während der empirischen Analyse spezifisch weiterzuentwickeln.[70] Wie anhand des vorangehenden Aufbaus dieser Arbeit ersichtlich ist, wird diesem Verständnis zur Forschungsmethodik gefolgt, weil innerhalb der Interviews als wesentliche Quelle für die Datenerhebung eine erkenntnisleitende, wenn auch ergebnisoffene Befragung einzelner Interviewpartner ermöglicht wird. Der Erkenntnisgewinn wird durch ein „induktiv-deduktives Wechselverhältnis“[71] erreicht, wobei theoretisch generiertes Vorwissen in Form von heuristisch-analytischen Fragenstellungen dialogisch in die Interviewsituation miteingeht.

Auf Basis der angestellten theoretischen Vorüberlegungen zu strategischen Prozessen des Wandels, findet die in Abschnitt 2.2.4 eingeführte Konzeptualisierung als Referenzrahmen auf die nachfolgende empirische Fallstudie Anwendung und diente hierbei in einer erkenntnisleitenden Funktion als Grundlage für die Entwicklung des Interviewleitfadens im Rahmen einer halbstrukturierten Befragungstechnik sowie das theoretische Kodieren[72] bei der Datenauswertung.

Um die Qualität[73] der empirischen Untersuchung zu optimieren, wird in der Literatur die Triangulation für die Datenerhebung[74] zur Vermeidung bzw. Verminderung von Verzerrungen empfohlen.[75] So wurden mehrere leitfadengestützte Interviews und eine Dokumentenanalyse zur Absicherung von Interpretationen und Ermöglichung sinnvoller Quantifizierungen durchgeführt. Weil Interviews immer eine subjektive Meinung darstellen, wird zur „Objektivierung“ weiterhin eine „Perspektiven-Triangulation“[76] angeregt, weswegen einzelne Interviews mit „Experten“[77] unterschiedlicher Positionen bzw. Rollen innerhalb der Veränderungsprozesse, unter Berücksichtigung individueller Sichtweisen auf den Untersuchungsgegenstand, geführt wurden.

Der entwickelte Leitfaden wurde für die Interviews einheitlich verwendet und ist dem Online-Appendix dieser Arbeit beigefügt (vgl. Online-Appendix 2). Der Leitfaden ist dabei nicht als starrer, chronologischer Ablaufplan zu verstehen, sondern dient im Wesentlichen als grobes Leitraster zur Orientierung über wichtige Forschungsfragen und Sicherstellung der Vergleichbarkeit, wobei Reihenfolge und Schwerpunktsetzung im Verlauf des Interviews situativ angepasst werden können.[78]

Um eine möglichst freie und ungestützte Darstellung der Eindrücke und Problemsicht einzelner Interviewpartner, ggf. über bereits angestellte theoretische Vorüberlegungen hinausgehende Erkenntnisse, zu erhalten, wurde zunächst zu allgemeinen Charakteristika der Geschäftsentwicklung sowie den persönlichen Einschätzungen zum spezifischen Verlauf und Zustandekommen von Veränderungsprozessen befragt und gegebenenfalls Rückfragen gestellt bzw. abschließend konkrete Themen angesprochen.

Die Suche nach potentiellen „Experten“ erfolgte durch fundierte Vorkenntnisse über relevante Akteure in ihrer spezifischen Rolle in Bezug auf die Geschäftstätigkeit im Marktsegment „Sanitär“ bzw. auf entsprechender Geschäftsbereichs-Ebene.[79] Hierbei wurde zwischen strategischen Entscheidungsträgern als „Führer“ sowie Betroffenen als „Geführte“ des Wandels unterschieden.

Da gemäß theoretischer Vorüberlegungen unter anderem auch politische und emotionale Sachverhalte für die Unter-

[67] Da die untersuchten Veränderungsprozesse in der Vergangenheit liegen, ist der Anspruch einer idealtypischen Untersuchung aktueller Ereignisse nicht erfüllt. Aufgrund der Bedeutung der Geschäftsaufgabe für den Bereich wird allerdings der Anspruch erhoben, dass die Qualität und Datentiefe der Informationen nicht durch den zeitlichen Abstand beeinträchtigt ist. In den Interviews hat sich gezeigt, dass sich alle Gesprächspartner sehr genau an die Ereignisse innerhalb des Betrachtungszeitraumes erinnern. Für Yin ist zum Gegenwartsbezug ausreichend, dass mit an den Vorgängen beteiligten Akteuren gesprochen wird (Yin, 2009:9f.).

[68] Vgl. zur Charakterisierung wesentlicher Formen und Merkmale qualitativer Interviews z. B. Lamnek, 2005:383.

[69] Vgl. Glaser und Strauss, 1965, 1967 als Gründungsväter sowie Weiterentwicklungen nach Glaser, 1978; Strauss, 1991 und Strauss, 1991; Strauss und Corbin, 1996.

[70] Vgl. Yin, 2009:14, 28ff.

[71] Witzel, 2000:2.

[72] Vgl. Strauss und Corbin, 1996:75-93.

[73] Vgl. als Überblick von Gütekriterien qualitativer Forschung z. B.Flick, 2005:487ff..

[74] Vgl. zur Daten-Triangulation z. B. Lamnek, 2005:159, Flick, 2005:330.

[75] Vgl. zur Triangulation Yin, 2009:116, Mayring, 2002:147f., Lamnek, 2005:158ff.. Vgl. zu Gütekriterien qualitativer Forschung ausführlich Wrona, 2005:39-46.

[76] Vgl. zur Perspektiven-Triangulation, die mitunter auch als Kategorie der Datentriangulation gesehen wird Flick, 2005:48ff., Flick, 1995:432.

[77] Die Interviews werden als besondere Form des Leitfadeninterviews im Sinne sog. „Experteninterviews“ abgehalten, wobei eine theoriegeleitete Auswahl der Interviewpartner zu entsprechenden Themenschwerpunkten innerhalb der Problemstellung erfolgt (vgl. z. B. Mayer, 2008:37f.).

[78] In Abhängigkeit der verschiedenen Rollen der Gesprächspartner und auf Basis der vorherigen Antworten wurden die offenen gestellten Fragen spezifisch angepasst und konkrete Themenschwerpunkte näher erfragt (vgl. zum Vorgehen halbstandardisierter Interviews mit Hilfe eines Leitfadens insb. Flick, 2005:118ff., Bortz und Döring, 2006:314, Yin, 2009:107, Mayring, 2002:67ff, Lamnek, 2005:329ff., Eisenhardt, 1989:539).

[79] Nach Zugang zum Feld durch bereits bestehende Kontakte in der Strategie-Abteilung des Geschäftsbereichs wurden im Rahmen der ersten Gespräche weitere Verbindungen zu potentiell relevanten Interviewpartner hergestellt. Zudem wurden einige Mitarbeiter im Feld zufällig befragt, um verschiedene Positionen und Funktionen gleichermaßen zu berücksichtigen und ggf. neue Sichtweisen zur Thematik zu integrieren.

suchung von Interesse sind und damit vor allem „weiche“ Faktoren angesprochen werden, wurde eine offene Gesprächsatmosphäre und hohe Vertrauensbasis zu den Betroffenen für eine tiefgreifende Ergründung des Phänomens geschaffen. Es wurde deshalb primär Wert darauf gelegt, dass der Interviewte aufrichtig sensible Fakten und persönliche Meinung zu nennen bereit ist und nicht der Eindruck entsteht, die Aussagen können im Nachhinein gegen ihn verwendet werden. Deshalb wurde bewusst auf eine Aufzeichnung der Interviews verzichtet. Es wurden handschriftliche Feldnotizen während der Interviews sowie unmittelbar danach angefertigt.[80] Die Einzelinterviews wurden persönlich im vertrauten Umfeld der Befragten, d. h. nach Möglichkeit in deren Büros am Firmenstandort durchgeführt und dauerten von fünfundvierzig Minuten bis zu zwei Stunden.

Tabelle 1 stellt die Auswahl der Personengruppen, die als Gesprächspartner fungierten, dar. Die Anzahl der Gesprächspartner ist jeweils in Klammern angegeben, die eingefärbten Interviewpartner beziehen sich auf ausgewählte Fälle[81] zur intensiveren Interpretation. Hierbei kam es zu Überschneidungen, da einzelne Mitarbeiter durch ihre berufliche Veränderung bzw. infolge Änderungen in der Organisationsstruktur während des Betrachtungszeitraums in verschiedenen Funktionen tätig waren. Insgesamt wurden vierundzwanzig Personen interviewt. Zwei Interviews wurden mit zwei Experten durchgeführt (I-I-04 und I-I-05), einige Experten wurden mehrmals befragt.

Aus den Interviews, wurden sechs Fälle bestimmt, die das Untersuchungsfeld in seiner Vielfalt möglichst breit aufspannen sowie bedeutsame Merkmalskombinationen hinreichend berücksichtigen und hierzu einer intensiven Interpretation unterzogen werden sollten.[83] Die Auswahl dieser sechs Fälle orientierte sich an ihrem Verhältnis zum Feld und folgte dem Prinzip der minimalen und maximalen Kontrastierung: So wurden drei Fälle der damals maßgebenden Entscheidungs- bzw. Wandlungsträger sowie drei Fälle von Betroffenen des Wandels ausgesucht. Alle Gesprächspartner haben dabei den gesamten Rückzugsprozess, teilweise in verschiedenen Positionen, verfolgt.

Da allen Interviewpartnern wegen der Sensibilität des Themas Vertraulichkeit zugesichert wurde, wurden die Interviews bei der Transkription anonymisiert. Die Transkripte wurden im Anschluss ausgewählten Interviewten zur Freigabe vorgelegt, um sicherzustellen, dass die Aussagen im Sinne einer „kommunikativen Validierung“[84] unverfälscht und wahrheitsgemäß wiedergegeben wurden.

Aufgrund der Umstände nicht vorliegender Tonaufnahmen der Gespräche basierte die Datenanalyse auf der Interpretation erfasster Textsequenzen der Interviewpartner sowie informellen Informationen, persönlichen Kommentaren, Einschätzungen oder spontanen Assoziationen.

Während des Kodierungsprozesses fiel auf, dass Aussagen über relevante Entscheidungen und Handlungen stark an bestimmte Ereignisse bzw. Veränderungen des internen oder externen Kontexts gekoppelt waren und widerholt hervorgehoben wurden. Um einzelne Veränderungen in chronologischer Zeitfolge nachzuvollziehen, wurden Textpassagen innerhalb der Kodierung zunächst mit einer zeitlichen Dimension versehen. Den Aussagen wurden dazu jeweils abzugrenzende Phasen im Rahmen des kontinuierlichen Wandlungsprozess zugewiesen, wobei zwischen Episoden[85] und Epochen[86] unterschieden wurde. Weiterführend erfolgte eine (interpretative) Textkodierung in Form kurzer Memos.

Die Kategorisierung lehnte dann stark an den vorstrukturierten Komponenten des eingeführten Bezugsrahmens (vgl. Abbildung 5) an. Vereinzelt prägten explizite Äußerungen der Gesprächspartner einzelne Subkategorien (im Sinne einer „In-Vivo“-Kodierung).[87] Das entstandene Kategoriensystem enthielt im Ergebnis drei identifizierte Kernkategorien mit jeweils mehreren Subkategorien (vgl. Online-Appendix 3).

Das entstandene Kategoriensystem enthielt im Ergebnis drei identifizierte Kernkategorien mit jeweils mehreren Subkategorien (vgl. Online-Appendix 3).

Um theoretische Implikationen in Bezug auf das Forschungsvorhaben abzuleiten, wurden identifizierte (Sub-)Kategorien der verschiedenen Fälle zu einander in Beziehung gesetzt und das Analyseraster in einem iterativen Prozess als Wechselspiel zwischen theoretischen Annahmen und empirischen Befunden gefestigt.

Ergänzt wurden die Interviews durch Dokumentenanalyse[88], wobei auf unternehmensinterne Archivbestände, wie Werbe-Broschüren, Dokumente zur Firmenhistorie, verfügbare Projektpräsentationen, Strategie- und Produktionsplanungsunterlagen, Memos, Briefe und E-Mails sowie Marktberichte, Fachzeitschriften und Zeitungsartikel zurückgegriffen wurde. Die verwendeten Dokumente liegen in Papierform vor und befinden sich in einem beigelegten Anlageordner dieser Arbeit.[89] Ein Überblick zur Datenbasis verwen-

[80] Die Aufzeichnungen wurden als kondensierte Darstellung in einem Gesprächsprotokoll zusammengeführt; wobei versucht wurde, wesentliche Inhalte möglichst unverfälscht und detailgetreu wiederzugeben bzw. zentrale Textsequenzen im Kontext als Zitate zu erfassen. Es wurde dabei eine tabellarische Form der Darstellung in Hinblick auf die anschließenden Schritte der Datenauswertung gewählt. Die Transkripte finden sich zusammen mit den Textkodierungen auf einer separaten Anlage-CD.

[81] Unter einem „Fall“ wird im Folgenden ein Interview verstanden (vgl. Wrona, 2005:28).

[83] Die Auswahl der Fälle für die intensive Analyse erfolgte in Kenntnis aller Interviews, sodass besondere Inhalte identifiziert und entsprechend relevante Fälle herausgegriffen werden konnten.

[84] Vgl. Scheele und Groeben, 1988.

[85] Im Sinne eines definierten Abschnitts der Veränderung, die neue Oberflächenstrukturen erzeugen (vgl. (Kutschker und Schmid, 2011:1110ff.).

[86] Ausdruck einer langfristigen Veränderung von Handlungsstrukturen oder bestimmten Paradigmen in Bezug auf Aspekte der Tiefenstruktur (vgl. Kutschker und Schmid, 2011:1112ff.).

[87] Vgl. Wrona, 2005:29.

[88] m Gegensatz zu den Erhebungsmethoden von Interviews und informelle Beobachtungen im Feld liegen Dokumente und Archivdaten bereits vor, sodass „nur bei der Auswahl der Dokumente, nicht aber bei der Erhebung [...] die Subjektivität des Forschers herein[spielt].“ Mayring, 1990:32).

[89] Die Dokumente wurden mit eigenen Signaturen versehen und bei der Kennzeichnung gemäß inhaltsanalytischer Kommunikationsmodelle (siehe u. a. Mayring, 2003:50) eine Unterscheidung von Quelle und Rezipient vorgenommen (vgl. Online-Appendix 4).

Tabelle 1: Datenbasis aus problemzentrierten Experteninterviews

Quelle	Ehemalige[82] bzw. aktuelle Position	Interview-Signatur
Betroffene des Wandels	Ehemaliger Manager Business Development PMMA-Halbzeuge (2x)	I-I-02 / I-I-23 / I-I-19
	Ehemaliger Marketingleiter PMMA-Halbzeuge (1x)	I-I-06 / I-I-24
	Director Business Development PMMA- Halbzeuge (1x)	I-I-08
	Marketingleiter PMMA-Halbzeuge (1x)	I-I-05
	Marktsegmentmanager PMMA-Möbel (1x)	I-I-05
	Marktforscher PMMA-Halbzeuge (1x)	I-I-12
	Manager Marktkommunikation PMMA (1x)	I-I-09
	Leiter Logistik/SCM PMMA (1x)	I-I-26
	Verkaufsleiter PMMA-Halbzeuge (1x)	I-I-16
	Ehemaliger Verkaufsleiter und Marktsegmentmanager Sanitär (1x)	I-I-18
	Produktmanager PMMA „Standard-Platten“ (1x)	I-I-15
	Ehemaliger Produktmanager PMMA-Platten Extrusion (2x)	I-I-04 / I-I-03 / I-I-14
	Ehemaliger Produktmanager PMMA-Platten Sanitär (2x)	I-I-03 / I-I-14 / I-I-13
	Ehemaliger Disponent PMMA-Platten (2x)	I-I-01 / I-I-26
	Ehemaliger Vorarbeiter PMMA-Platten Guss (2x)	I-I-07
	Ehemaliger Planungsingenieur „Gießlinie 3“ (1x)	I-I-21
	Ehemaliger Betriebsingenieur Gussproduktion (1x)	I-I-02 / I-I-23
	Ehemaliger Leiter Technikum Extrusion(1x)	I-I-22
	Ehemaliger Betriebsleiter PMMA-Platten Extrusion (3x)	I-I-02 / I-I-23 / I-I-22 / I-I-11 / I-I-20
	Ehemaliger Betriebsleiter PMMA-Platten Guss (1x)	I-I-02 / I-I-23
	Ehemaliger Produktionskoordinator PMMA-Platten Extrusion (1x)	I-I-25
	Ehemaliger Controller PMMA-Platten (2x)	I-I-06 / I-I-24 / I-I-17
Entscheidungs-/ Wandlungsträger	Ehemaliger Betriebsleiter PMMA-Platten Guss(1x)	I-I-21
	Ehemaliger Produktionskoordinator PMMA-Platten Guss (1x)	I-I-04 / I-I-10
	Leiter Produktion & Technik PMMA (1x)	I-I-11 / I-I-20
	Bereichsleiter PMMA (1x)	I-I-06 / I-I-24

deter Dokumente und Archivmaterialien verschafft Online-Appendix 4. Durch die Analyse hinzugezogener Dokumente konnten einzelne Entwicklungen und Zusammenhänge, welche nicht im Rahmen der Interviews erschlossen werden konnten bzw. sich im Detail erheben ließen, im Sinne einer verlässlichen Datengenerierung genauer nachvollzogen und interpretiert werden (vgl. Mayring, 1990:32f.). Hinsichtlich der Zuverlässigkeit der Datenlage ist hierbei kritisch anzumerken, dass infolge von Umstellungen der IT-Systeme vor allem Controlling-Daten vor dem Jahr 1996 nicht lückenlos vorhanden waren und sich besonders die Suche nach Archivmaterialien zu strategischen Entscheidungsprozessen unter verschiedenen Führungswechseln und dem Ausscheiden wichtiger Entscheidungsträger herausfordernd gestaltete.

3.2. Ausgangslage: Die PlastiCo GmbH und das PMMA-Halbzeuge-Geschäft

3.2.1. Historische Entwicklung der PlastiCo GmbH als Teil der ChemCo AG

Die in Deutschland ansässige ChemCo ist ein weltweit führender, im September 2007 unter diesem Namen etablierter Konzern der Spezialchemie. Das börsennotierte Unternehmen beschäftigte Ende 2013 rund 33.000 Mitarbeiter und wies bei einem Umsatz von 12,9 Mrd. Euro ein operatives Ergebnis (bereinigtes EBITDA) von 2 Mrd. Euro aus.[90]

Eine der historischen Wurzeln des Konzerns reicht in das Jahr 1907 zurück, als die PlastiCo OHG in Esslingen gegrün-

[90] Siehe zur Unternehmensdarstellung der ChemCo ausführlich ChemCo, 2014a.

det wurde.[91] Nachdem das Unternehmen, das zunächst Hilfsmittel für die Lederindustrie und Waschmittel auf Enzymbasis herstellte, seinen Werksstandort zur Produktionsausweitung nach Darmstadt verlagert hatte, gelang mit der Erfindung des bruchfesten, glasklar transparenten Kunststoffs Polymethylmethacrylat (PMMA) in den 1930er Jahren der endgültige Durchbruch für das Unternehmen, als fortan marktführender Anbieter von Produkten der Methacrylchemie.

Von nahezu Anbeginn an als global operierendes Unternehmen[92], bei stetigem Ausbau der Produktion[93] sowie Entwicklung und erfolgreicher Vermarktung neuer Chemieprodukte[94] zur Diversifikation[95], wurde die Firma ab 1938 als GmbH geführt und erhielt mit dem Ausscheiden einer der beiden Gründerfamilien aus dem Gesellschafterkreis im Jahre 1971 den alleinigen Firmennamen der Gründerfamilie.

Nachdem die MineriCo AG ab den 1980er Jahren die PlastiCo GmbH sowie die Chemische Fabrik SapoCo GmbH übernommen hatte, erfolgte nach einer Phase der Unternehmensrestrukturierung 1999 zunächst die Verschmelzung mit der AuriCo AG zur AuriCo-MineriCo AG. Unter Bündelung der Kompetenzen von insgesamt fünf Vorgänger-Chemieunternehmen fusionierte die AuriCo-MineriCo zusammen mit der FerrensiCo schließlich wieder, zur alleinigen „AuriCo AG“. Im Zuge der „Konzentration auf Spezialchemie“ wurden zahlreiche, nicht mehr dem Kerngeschäft zugehörige Aktivitäten veräußert und die operativen Methacrylatchemie-Aktivitäten auf drei unterschiedliche Geschäftsgebiete verteilt. Im Rahmen einer weiteren Übernahme durch den CarbiCo-Konzern und Umstrukturierung firmierte der Bereich mit Schwerpunkt Spezialchemie 2007 schließlich unter der heute börsennotierten ChemCo AG (vgl. Abbildung 6).

Das traditionelle Methacryl-Geschäft und der ehemals eigenständige PMMA-Geschäftsbereich der PlastiCo ist heute innerhalb eines Segments einer eigenständigen Geschäftseinheit der ChemCo integriert.[96]

[91] Siehe zur Historie der PlastiCo GmbH ausführlich ChemCo, 2014b; GmbH, 2007.

[92] Bei zunächst systematischem Aufbau des Vertriebs in Deutschland, wurden 1909 Vertretungen im französischen Lyon und in den USA, in Philadelphia eingerichtet. Auslandsgeschäfte zur Produktion von PMMA in Spanien (PlastiCo S.A.) und Italien (PlastiCo S.p.a.) erfolgten ab den 1950er Jahren.

[93] Die Inbetriebnahme eines Rohstoffwerks in Worms zur Herstellung von MMA, u. a. Vorprodukt von PMMA, erfolgte 1955. Ein zweiter Produktionsstandort zum Ausbau der Produktionskapazitäten von Platten aus PMMA wurde 1969 in Weiterstadt eröffnet. Ab 1968 wurden dort der Extusionsbetrieb sowie Hallen für Plattenzuschnitt, Lagerung und Versand in Betrieb genommen. Die Aufnahme der Gussproduktion erfolgte 1971.

[94] u. a. VISCOPLEX Öladditive, ROHACELL Hartschaumstoffe und Arznei-Überzug zur gesteuerten Freisetzung von Wirkstoffen unter der Markenbezeichnung EUDRAGIT

[95] Der Begriff ist hierbei allein auf die Leistungsbreite durch die Anzahl und der Verbundenheit von Produkten und Märkten bezogen. Strategische Verhaltensweisen einer vertikalen bzw. geographischen Dimension werden nachfolgend in Form von Strategien zur vertikalen Integration bzw. Internationalisierung betrachtet (vgl. zur Diskussion der Begriffe ausführlich Bamberger und Wrona, 2012a:153ff.).

[96] Vgl. Online-Appendix 6 zur aktuellen Struktur und Organisation und ausführlich ChemCo, 2014d.

3.2.2. PMMA-Produkt-/Marktaktivitäten

Als Vollanbieter von PMMA-Halbzeugen (PMMA-Platten, - Rohre, - Stäbe, -Folien) und -Formmassen[97] wurde ab den 1930er Jahren bis heute eine breite Produktfamilie geschaffen, die in vielen Industriebranchen, wie der Architektur, Lichtwerbung und Leuchten-Industrie, im Transportwesen oder im Sanitärbereich, Verwendung fand.[98] Die Aufteilung des heterogenen Gesamtmarkts für Halbzeuge aus PMMA und Definition einzelner strategischer Geschäfte erfolgt danach anhand der unterschiedlichen Anwendungsbereiche und Ausrichtung an Käufergruppen (Marktsegmentierung, vgl. auch Abbildung 8).

PMMA-Produkte werden traditionell, unter Verwendung zweier unterschiedlicher Herstellungs-Technologien, als „gegossener“ (PMMA GS) und/oder „extrudierter“ (PMMA XT) Produkt-Typ in zwei verschiedenen Produktions-Betrieben (Guss-Betrieb und Extrusions-Betrieb) produziert. Durch das Mischen von flüssigem MMA zwischen zwei Glasplatten im sog. Flachkammerverfahren auf Gießlinien und anschließende Aushärteprozesse (Polymerisation, u. a. durch Wasserbad-, Luftschrank-Polymeristation oder im Rostero-Verfahren) entstehen vernetzte, hochmolekulare, „gegossene“ Acrylglas-Platten, auch Acrylguss genannt. „Extrudierte“, nieder- bis höhermolekulare PMMA-Platten werden hingegen aus bereits polymerisiertem MMA (PMMA-Formmasse) hergestellt. Dieses wird in Extrudern aufgeschmolzen und durch entsprechende Formgebungseinrichtungen zu Massivplatten, sowie Steg- oder Wellplatten, Stäben und Rohren stranggepresst.[99]

Die beiden chemisch unterschiedlichen Produkt-Typen unterscheiden sich teilweise hinsichtlich ihrer anwendungsbezogenen Charakteristik und werden entsprechend verschieden eingesetzt.[100] In Bezug auf ihre technischen Produkteigenschaften und der Wahrnehmung beim Endkunden sind die beiden Typen allerdings weitestgehend vergleichbar, sodass von einer Produktähnlichkeit auszugehen ist. An dieser Stelle sei auf die hohe Produktverbundenheit innerhalb der PMMA-Familie über die Input-Output-Beziehung hingewiesen: Die eingesetzten Vorprodukte (MMA bzw. PMMA-Formmasse) zur Herstellung von PMMA-Halbzeugen (gegossen bzw. extrudiert) werden für die interne Weiterverarbeitung selbst produziert sowie von einem eigenständigen Geschäftsgebiet am Markt vertrieben. Der Vertrieb von Halbzeugen aus PMMA erfolgt überwiegend indirekt, vereinzelt bestehen Direktkundengeschäfte.

[97] Perlförmige Kunststoffgranulate zur Weiterverarbeitung, v. a. im Spritzgussbereich und für Extrusionsanwendungen (vgl. Keim, 2006).

[98] Siehe zu aktuellen Acrylglas-Anwendungen D-ie-07.

[99] Vgl. zur Herstellung und technischen Daten der beiden Produktgruppen ChemCo AG, 2013.

[100] „Gegossenes“ PMMA zeichnet sich besonders durch eine hohe Oberflächenqualität in breitem Verarbeitungsspielraum gegenüber extrudierten Sorten aus. Extrudiertes Acrylglas bietet Vorteile einer herstellungsbedingt geringeren Dickentoleranz sowie einfacheren, wirtschaftlichen Verarbeitung durch kürzere Umformzyklen bei geringeren Temperaturen (vgl. zu den Anwendungsbereichen der jeweiligen Produktgruppen ChemCo, 2014c).

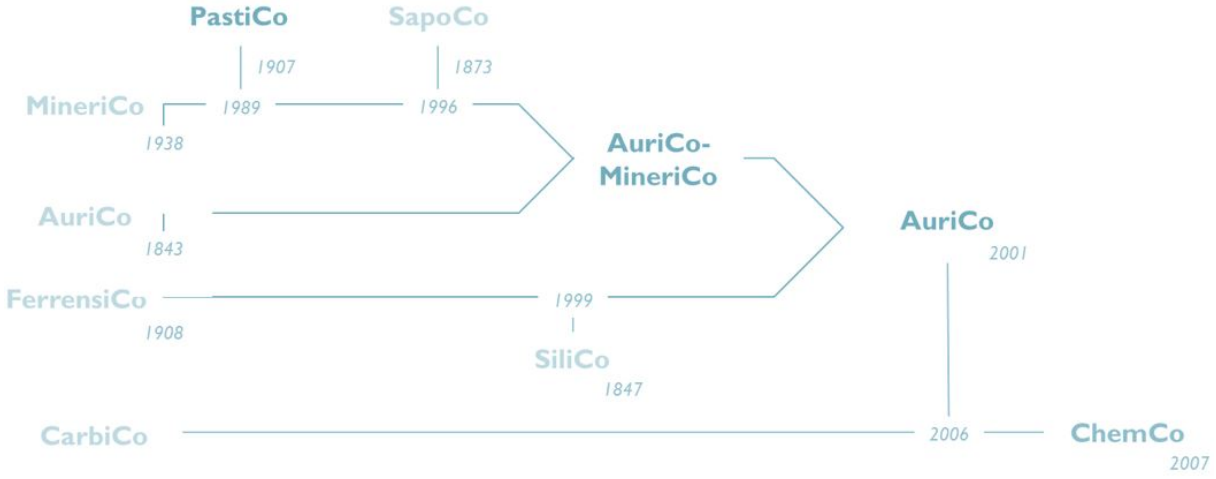

Abbildung 6: Vorgängergesellschaften von ChemCo (Quelle: GmbH, 2007).

3.2.3. Begründung des PMMA-Sanitär-Geschäfts und Marktentwicklung

Ein bedeutendes Anwendungsgebiet von PMMA entwickelte sich mit der Nutzung gegossener PMMA-Platten im Sanitärbereich, zur Herstellung von Bade- und Duschwannen.[101]

Der Sanitär-Markt für PMMA entstand dabei ursprünglich in England, als Ende der 1950er Jahre Acrylguss-Wannen zwecks Gewichtseinsparung gegenüber traditionellen Modellen aus Stahl-Emaille für Übersee-Exporte gefertigt wurden (vgl. D-ee-01). Etwa zur selben Zeit beschäftigten sich Ingenieure bei PlastiCo mit weiteren zivilen Anwendungsgebieten des marktetablierten PMMA neben einer bereits weiten Verwendung in Lichtwerbung, bei Dachverglasungen und in der Fassadengestaltung. Die sehr gute Bruchfestigkeit, lange Lebensdauer sowie insbesondere leichte Bearbeitbarkeit und damit auszeichnende hohe Designfreiheit inspirierte zu neuen Form- und Farbvarianten von Badewannen durch den Einsatz von Acrylguss-Platten als Grundwerkstoff zur Verarbeitung zu Badewannen im Vergleich zu derzeit herkömmlichen Modell-Standards aus emaillierten Stahl und Keramik. Da verarbeitende Hersteller von Bade- und Duschwannen zunächst nicht an den Erfolg von Acrylwannen glaubten, wurden die Wannen vorerst vollständig in Eigenregie im Werk Darmstadt hergestellt und vertrieben und damit das „Sanitärgeschäft“ für Acrylguss begründet. Im Zuge eines sich abzeichnenden Markttrends zu Acryl-Designobjekten bei aufgenommenen Aktivitäten zur Vermarktung, u. a. durch Vorstellung des PMMA-„Wohnbades“ auf Fach- und internationalen Kunststoffmessen[102] Ende der 60er, wurde großindustriell gefertigt und PlastiCo zog aus dem Fertigteilsektor zurück. PlastiCo konzentrierte sich als traditioneller Acrylglas-Hersteller und B2B-Lieferant wieder ausschließlich auf die Produktion von Acrylglas-Platten in vielfältigen, sanitärkundenspezifischen Farbvarianten (vgl. zu PMMA GS SW-Sorten D-ii-02) und Plattenformaten (Sanitärproduktion als ein „Zuschnitt-Geschäft“) und baute ein Direktkundengeschäft zu dreistufig operierenden Wannenproduzenten in Europa auf (vgl. zur Wertschöpfungskette Abbildung 7, zur Einordnung von Sanitär-Produkt-/Marktkombinationen innerhalb von PMMA Produkt-/Marktaktivitäten Abbildung 8).

Die volumenmäßig bedeutendsten Märkte für Acrylguss-Platten zur Herstellung von Badewannen, Whirlpools und Duschwannen sowie später auch Wannenschürzen und Wänden von Duschkabinen entwickelten sich in den 1970er Jahren in Europa, im nahen und mittlere Osten, Asien, Australien, Sudafrika sowie in den USA. Neben einer Vielzahl kleinerer Halbzeuge-Hersteller, teilweise nur von regionaler Bedeutung, dominierten neben PlastiCo, fortan zwei weitere, weltweit führende Kunststoffproduzenten (PolyCo und ResiCo) den europäischen Markt.[103]

Der Sanitär-Markt stellte das zu diesem Zeitpunkt größte Marktsegment[104] von PlastiCo für Produkte aus PMMA dar. Produziert wurde im Verlauf des unternehmerischen Wandels an verschiedenen Standorten, ein Überblick zu historischen Produktionsstandorten für Sanitärmaterial von PlastiCo sowie der beiden anderen Marktführer verschafft Abbildung 9.

Im Verlauf der Marktreifung wurden internationale Produktnormen zur Spezifizierung von Acrylplatten eingeführt

[101] Die PMMA-Platten werden dazu im Tiefziehprozess durch Erwärmung und angelegten Unterdruck in Form gezogen und der Wannenrohling zwecks Versteifung mit Glasfaserverbundklebstoff besprüht sowie abschließend an den Rändern zugeschnitten und mit entsprechenden Bohrungen für Ab- und Überlauflöcher versehen (siehe zur Herstellung von Acrylwannen D-ee-02).

[102] Vgl. die Vorstellung des Acrylglas-„Wohnbades“ auf der internationalen Kunststoffmesse „K67“ in Düsseldorf sowie Fachmesse „Bau 68“ in München (Spiegel, 1968).

[103] Vgl. zur prozentualen Aufteilung des Marktes unter den Herstellern Online-Appendix 7.

[104] Die Bedeutung des Sanitärgeschäfts als Haupt-Absatzsegment für Halbzeuge aus PMMA wurde in den Interviews mehrfach hervorgehoben (vgl. #141, #226, #286). Aufgrund der großen Verkaufsmengen waren mit einer alleinigen Desinvestition des Geschäfts umfangreiche Veränderungen in Bezug auf grundlegende Prozesse und Strukturen innerhalb des gesamten Geschäfts mit PMMA verbunden.

Abbildung 7: Wertschöpfungskette gegossene Acryl-Wannen nach dreistufigen Vertriebsweg (Im Zuge der Marktetablierung sah sich PlastiCo als rückstufig integrierter Zulieferer gegossener Acrylguss-Platten (Rohstoffproduzent von MMA und Verarbeitung zu Acrylguss-Platten, in „lila“ dargestellt). Die Verarbeitung zu Bade- und Duschwannen erfolgte durch Wannenhersteller, welche die Fertigprodukte auf dreistufigem Vertriebswege an den Endkunden liefern (in grau-Abstufungen dargestellt).) (Quelle: eigene Darstellung)

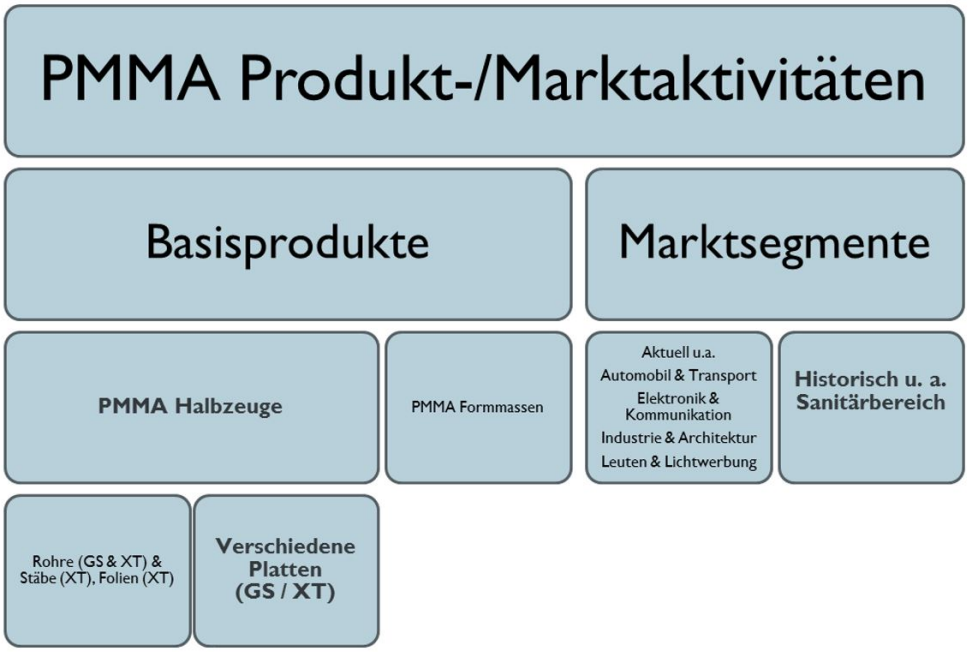

Abbildung 8: Übersicht PMMA Produkt-/Marktaktivitäten und Einordnung des „Sanitär“-Geschäfts (Quelle: eigene Darstellung).

und innerhalb Europas festgelegte Prüfverfahren zur Gewährleistung des Qualitätsstandards vernetzt gegossener Acrylplatten zu Anwendung für Bade- und Duschwannen etabliert (EN 263, ASTM (ASTM, 2002) D4802, EN 198, EN 249, ANSI-Z-124.1 (ANSI, 2005)).

Die Wettbewerbssituation in Europa verschärfte sich im Zuge der Commodisierung von Acrylguss unter Eintritt neuer Wettbewerber und Importe aus Ländern geringerer Herstellkosten (durch hohe Lohnkostendifferenzen) und spiegelte sich ab den 1990er Jahren in einem sukzessiven Preisverfall wieder. Aufkommende Substituts-Werkstoffe beeinflussten die Branchenattraktivität weiterhin negativ, das Angebot preisniveaugleicher Fertigteile von Acrylwannen führte ab 2002 zur Erosion des 3-stufigen Vertriebsweges (siehe zu Marktgesprächen zum Direktbezug D-ee-03, D-ee-04). Der qualitative Verlauf der Marktentwicklung ist in Abbildung 10 nachgezeichnet.

Mit der Eingliederung des PMMA-Geschäfts von PlastiCo in die Konzernstrukturen der MineriCo AG wurde das Sanitärgeschäft bei zuvor regionaler Betreuung zentralisiert und hinsichtlich einer herausgestellt unbefriedigenden Geschäftsperformance eine Rückzugsstrategie verfolgt. Bei einem zunächst selektiv vollzogenem Rückzug im Jahr 1996 durch Fokussierung auf bestimmte Kunden und Produkte, erfolgte unter Führung der AuriCo im Geschäftsjahr 2008 der vollständige Rückzug aus dem Sanitärgeschäft.

3.3. Theoretische Analyse des strategischen Wandels

3.3.1. Objektbereich und Art des Wandels

Wie bereits dargestellt, bezieht sich das zu untersuchende Wandel-Phänomen auf die strategischen Veränderungsprozesse zum vollständigen Ausstieg aus dem Sanitärgeschäft.

Die realisierte Rückzugsstrategie auf Geschäftsfeldebene ist zunächst als Form von Geschäftsentwicklungsstrategien zu sehen und liegt in den Merkmalen des Marktes im Rahmen der natürlichen Marktreifung und dabei unbefriedigenden Entwicklung der Geschäfts-Performance begründet.[105] Damalige Akteure sprechen z. B. von einem „marktgetriebenen“ (#53, #230) Rückzug und betonen dahingehend ein re-

[105] Vgl. #15, #24, #62, #90, #101, #159, #226, #301, #439f..

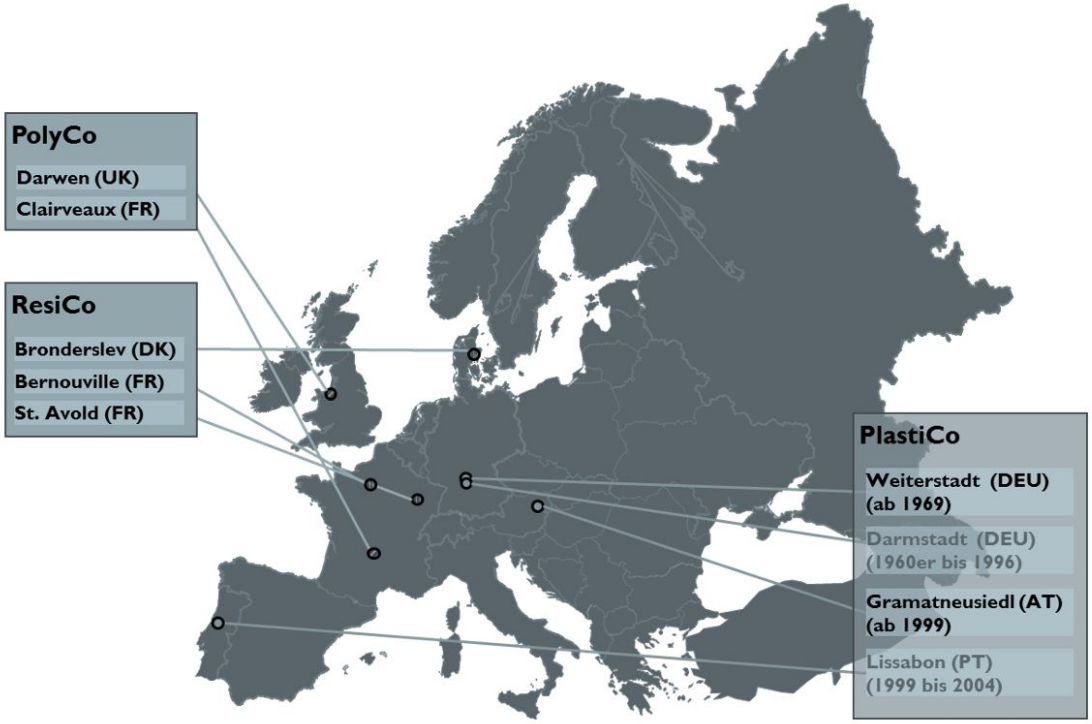

Abbildung 9: Überblick historischer Produktionsstandorte führender Hersteller von Sanitär-Acrylgussplatten (Quelle: eigene Darstellung).

aktives Entscheidungsverhalten bei zunehmender Commodisierung von Acrylguss und einem vorherrschenden Verdrängungswettbewerb.

Als Dimension von Wettbewerbsstrategien bildet die Desinvestition gleichsam eine spezifische, dynamische Form der Refokussierung des PMMA-Geschäfts sowie Aktivität zur strategischen Neuausrichtung bei Etablierung eines neuen Kerngeschäfts. Diese Zusammenhänge gilt es nachfolgend noch genauer zu erörtern.

Der Marktrückzug erfolgte sukzessive, im Rahmen verschiedener Prozesse von Entscheidungen, Handlungen und Interaktionen auf mittlerer Managementebene[106] sowie Veränderungen des strukturellen Rahmens, über einen längeren Zeitraum von 13 Jahren. Das folgende Statement eines maßgebenden Entscheidungsträgers charakterisiert den spezifischen Verlauf des Rückzugsprozesses:

> *"Der Rückzug aus Sanitär, das waren mehrere Schritte." (Geschäftsbereichsleiter #99)*

Als *„schrittweise"* (#99, #76ff., #327) vollzogenen, *„langsamen"* (#488) und *„schleichenden"* (#311) Rückzug beschreiben auch Betroffene der unteren Führungsebene die Geschäftsaufgabe und heben hierbei spezifische Episoden der Entscheidungsfindung- und Umsetzung, teils auch mit Unterstützung durch externe Berater und Anwendung speziell entwickelter Methoden, innerhalb des organisationalen Geschehens heraus.

Der Versuch einer ganzheitlichen Beschreibung der hierbei komplexen Ursache-Wirkungs-Beziehungen, insbesondere hinsichtlich zentral kognitiver Analysedimensionen, würde durch verschiedene Limitationen des angelegten Fallstudiendesigns spekulative Züge tragen. Wie im Theorieteil erläutert, schließen Prozesse organisationalen und strategischen Wandels vielfältige Phänomene ein, die in ihrem wechselseitigen Zusammenwirken grundsätzlich schwer zu erfassen sind. Dennoch soll das Zustandekommen des Marktrückzuges innerhalb des vollzogenen strategischen Wandels des Halbzeuge-Geschäfts mit PMMA bei einer mit dieser Arbeit verfolgten dynamischen Perspektive von Strategieprozessen, die einen kontinuierlichen organisatorischen Wandel integrieren, reflektiert und Implikationen hinsichtlich zugrundeliegender Ursachenkomplexe abgleitet werden.

3.3.2. Verlaufsform des Wandels: Episoden und Triebfedern

Bezogen auf den Betrachtungszeitraum des fokussierten Rückzugsprozesses illustriert Abbildung 11 zentrale Befunde der empirischen Analyse zur Beschreibung des vollzogenen Wandels und Prozessen des Zustandekommens. Es ist die Geschäftstätigkeit im Sanitär-Marktsegment, repräsentiert durch die jährlich abgesetzten Produktmengen [Jato=Jahrestonnen] auf der vertikalen Primärachse, kom-

[106] Als dritte Führungsebene (nach Vorstand/Geschäftsführung und Senior Management) und darunter: die Führungsebene, die direkte Mitarbeiterführung innehat und politische Entscheidungen umsetzt.

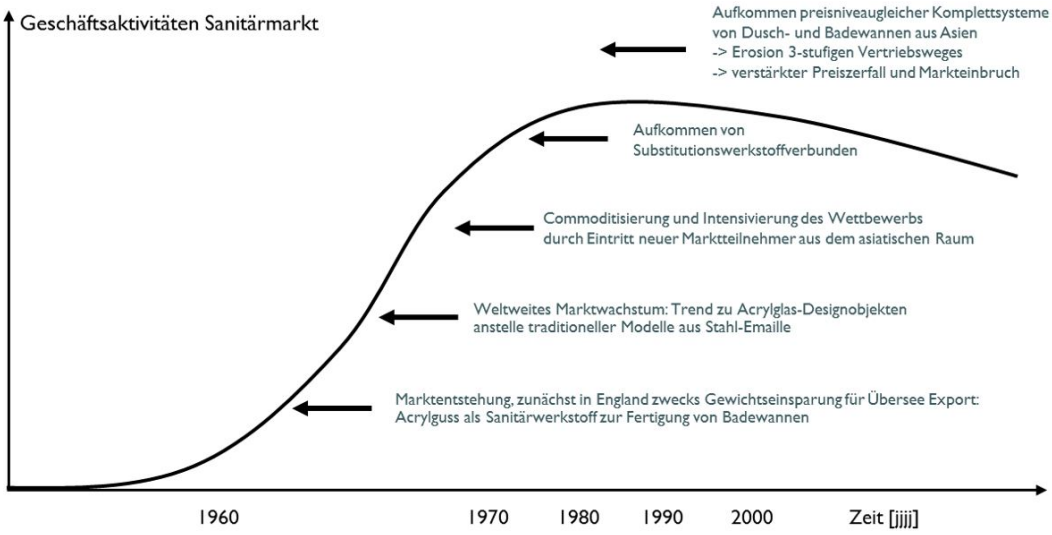

Abbildung 10: Überblick historischer Produktionsstandorte führender Hersteller von Sanitär-Acrylgussplatten (Quelle: eigene Darstellung).

biniert mit Informationen spezifischer Prozessmerkmale von unternommenen Aktivitäten zum Wandel (A), unternehmensinternen (I) und unternehmensexternen Ereignisse bzw. Entwicklungen (E) auf einer vertikalen Sekundärachse, dargestellt.

Obgleich das zweidimensionale Schaubild verschiedene Arten von Informationen enthält, sind die einzelnen Veränderungen hierbei nicht isoliert voneinander zu sehen, sie finden vielmehr parallel und integriert in einem gemeinsamen Entwicklungsprozess statt. Von besonderem Interesse sind die kausalen Zusammenhänge zwischen einzelnen Veränderungen. Die Veränderungen innerhalb der farbig hervorgehobenen Geschäftsjahren 1996, 2003 und 2008 werden im Verlauf der Analyse besonders herausgestellt.

Nachfolgend wird der historische Entwicklungsverlauf nachgezeichnet und auf die graphische Verdeutlichung hierbei zentraler Ereignisse bzw. Entwicklungen unter Verwendung entsprechender Bezugssymbole (A / I / E) verwiesen werden.

1907-1989 Familienunternehmen PlastiCo: Ende einer Ära. Nachdem das mittelständische[107], traditionell familiengeführte Unternehmen PlastiCo mit der Produktentwicklung von PMMA bereits vielfältige Anwendungsgebiete für Acrylglas erschlossen und sich dabei ein Image als Marktführer und Vollanbieter für PMMA-Halbzeuge aufgebaut hatte[108], veränderten sich die Markt- und Wettbewerbsbedingungen in den 70er und 80er Jahren grundlegend, im Rahmen eines Wandels von einem Verteiler- zu einem Verdrängungsmarkt.[109] Dabei planten die Geschäftsführer eine langfristige Kapazitätserweiterung im Bereich der Gussproduktion und Vertiefung der Technologie zur Serienfertigung von PMMA-Platten bei angestrebter Vollautomatisierung. 1986 wurde innerhalb eines Cross-licensing Agreement mit einem technologieführenden Wettbewerber[110] eine spezifische Investitionsentscheidung für den Bau einer entsprechend „hochproduktiven Anlage“ (#364) getroffen (A). In Form einer dritten Gießlinie „Gießlinie 3“ sollten damit die Produktionskapazitäten für „Standardguss-Produkte“[111] am Standort in Weiterstadt ausgebaut werden:[112]

> ”Die Enkel des Erfinders, man hat das Glänzen noch in den Augen gehabt. Es wurden Gefühlsentscheidungen getroffen.” (Geschäftsbereichsleiter #184f.)

1989-1999 Übernahme MineriCo: Refokussierung und Kostenmanagement. Grundlegende organisationale Änderungen erfolgten mit dem Aufgang der PlastiCo als eigene Legaleinheit im Chemiekonzern und der Übergabe der Geschäftsführung an das familienfremde Management der MineriCo AG (I). Der heutige Bereichsleiter des Geschäfts mit PMMA berichtet von einer zunehmenden Formalisierung strategischer Entscheidungsprozesse unter Abhaltung regelmäßiger „Strategiemeetings“ (#183) und größeren Distanz der neuen Geschäftsführung zum operativen Geschäft mit PMMA. Ehemalige Betriebsleiter beschreiben ein fortan erhöhtes Ergebnis- und Kostenbewusstsein bei Einführung

[107] In der Literatur konnte sich keine einheitliche Definition des mittelständischen Unternehmens durchsetzen. Zur quantitativen Abgrenzung vgl. Institut für Mittelstandsforschung, 2014, zur qualitativen Einordnung siehe etwa Roseboom, 2004:11f., ähnlich Geiser, 1983.

[108] Vgl. #28, #120, #322, #373.

[109] Vgl. #55, #179, #212.

[110] Das Partnerunternehmen erhielt im Gegenzug die Nutzungsrechte an der Extrusionstechnologie zur Herstellung extrudierter Acrylglas-Platten.

[111] Im Sinne universaler Produkte ohne anwendungsspezifische Eigenschaften in Abgrenzung zu sog. „Spezialitäten“. „Spezialitäten“ zeichnen sich dadurch aus, dass der Kunde bei vorhandenen Wechselbarrieren aufgrund von Alleinstellungsmerkmalen einen höheren Preis zu zahlen bereit ist bzw. der Preis nicht das dominante Kaufkriterium ist. Das Sanitärgeschäft ist hinsichtlich der Margenentwicklung als „Standardgeschäft“ anzusehen (D-ei-01:6).

[112] Vgl. #333, #354, #364, #442ff..

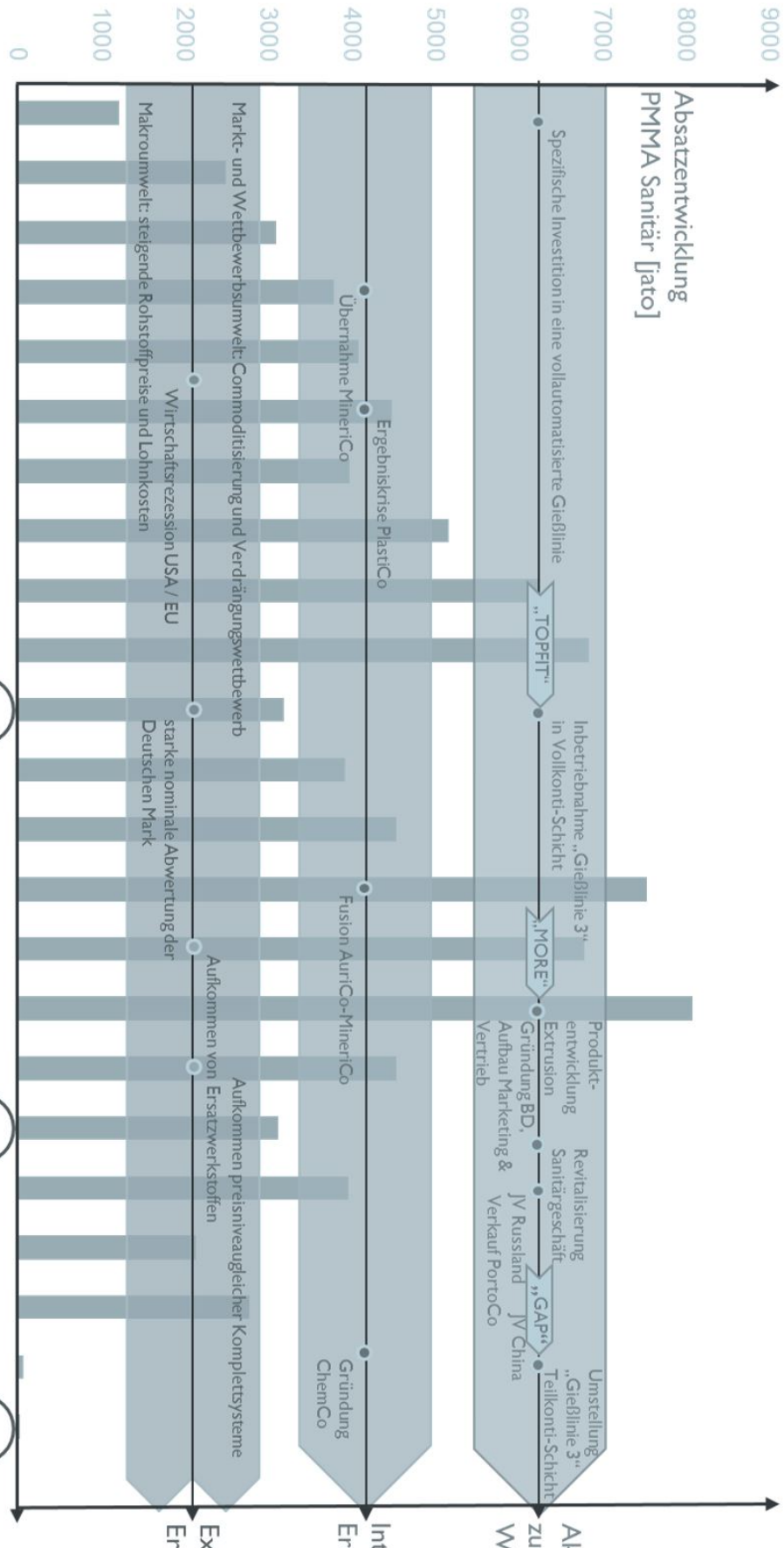

Abbildung 11: Überblick historischer Produktionsstandorte führender Hersteller von Sanitär-Acrylgussplatten (Quelle: eigene Darstellung).

von Ergebnis- und Marktsegmentrechnungen im Bereich der „Perle der Methacrylatchemie“ (#399) sowie verändertes Geschäftsverständnis und „strategisches Denken“ (#459) des Topmanagements.[113] **1994-1996 „TOPFIT“.** Bei drastischer Wirtschaftsrezession in den USA sowie Westeuropa (E), steigender Rohstoffpreise (E) und einem verschärften Wettbewerb (E) hatte PlastiCo Anfang der 1990er Jahre rapide Umsatzeinbrüche zu verzeichnen. Als daraufhin die Ertragskrise (I) offenbar wurde, erfolgte zwischen 1994 und 1996 eine umfassende Unternehmensrestrukturierung zur „Effizienzsteigerung und Konzentration auf das Kerngeschäft“ (GmbH, 2007:96). Mit externer Beraterleistung von McKinsey[114] wurden im Rahmen des Projekts „TOPFIT“ (A) einzelne Geschäftsgebiete ausgegliedert oder verkauft und Einsparungs- bzw. Optimierungspotenzial bestehender Produktions- und Arbeitsprozesse aufgezeigt. Betroffene erinnern sich an eine „40 %-Hürde“ (#6, #338) zur Senkung beeinflussbarer Kosten und dadurch massive Personalfreisetzungen. Im Zuge einer Refokussierung des PMMA-Geschäfts wurden ein Produktionsstandort im Ausland geschlossen, der Guss-Betrieb in Darmstadt stillgelegt und die Platten-Produktion von PlastiCo hierbei vollständig an den Standort nach Weiterstadt verlagert.[115]

Bei der Optimierung des PMMA-Produktportfolios wurde dabei im Wesentlichen der Rückzug aus dem Direktkunden-Geschäft des Sanitär-Marktsegment beabsichtigt, das sich bei der Analyse des als stark unprofitabel herausgestellt hatte (A).[116] Aus Sicht relevanter Management-, Controlling- und Steuerungsprozesse war innerhalb der Implementierung die Entwicklung der Kundenergebnisrechnung maßgebend, auf deren Basis unprofitable Sanitär-Kunden durch Preiserhöhungen gemanagt und die entsprechend kundenspezifischen Produktvarianten aufgegeben wurden.[117] In Form eines **teilweise realisierten Rückzugs aus dem Sanitär-Markt** erfolgte 1996 eine deutliche Reduzierung der Absatzmengen und Straffung des Produktportfolios.

Der Bereichsleiter spricht von der „ersten Phase der Konsolidierung“ (#104) und einem bezeichnenden „Paradigmenwechsel“ (#191) durch die Restrukturierung mit McKinsey, wobei die strategische Neuausrichtung zeitlich nachgelagert war:

> texit”Die erste Entscheidung [zum Rückzug aus dem Sanitärgeschäft] war kein strategisches Projekt. Die strategische Neuausrichtung erfolgte erst sieben bis acht Jahre später.” (Geschäftsbereichsleiter #151)

„Gießlinie 3“. In den darauffolgenden Jahren wurden die Produktaktivitäten im Sanitär-Segment allerdings wieder erhöht und Verkaufsmengen bei negativen Ertragsraten bis zu einem maximalen Absatz im Jahre 2001 gesteigert. Damalige Controller erklären die positive Mengenentwicklung durch die Produktionsumstellung auf die zeitgleich der Refokussierung von 1996, neu in Betrieb gehende „Gießlinie 3“ (A). Hinsichtlich einer veränderten Nachfragesituation und dabei wider Erwarten stagnierenden Wachstums für Acrylgussplatten bestanden hohe Überkapazitäten auf der Gießlinie bei vollkontinuierlicher Betriebsweise im Vierschichtbetrieb. Gemäß einer „opportunistischen Denkweise“(#161) des operativen Produktionsmanagements wurden Produktgruppen der übrigen Gießlinien nach verfahrenstechnischer Möglichkeit auf die vergleichsweise produktivere „Gießlinie 3“ verlagert. Die Sanitär-Produktion wurde - in Anbetracht freier Produktionskapazitäten - als Deckungsbeitragsgeschäft auf „Gießlinie 3“ weiter fortgesetzt. Die weniger produktionseffiziente „Gießlinie 2“ wurde zusammenhängend stillgelegt.[118]

> *”Die Absatzzahlen sollten hoch, das Problem war, die Absätze gingen nicht hoch. Dann zog man immer mehr Produkte auf Gießlinie 3, um sie in den Vier-Schicht-Betrieb zu bekommen (#12)... Es war eine produktionsgetriebene Entscheidung.” (Produktionskoordinator Gussbereich #18)*

Commoditisierung und Margenverlust. Unterdessen veränderte sich das Wettbewerbsumfeld des Konzerns in Märkten von PMMA-Anwendungen. Durch einen verstärkten Verdrängungswettbewerb und steigende Überkapazitäten seit der Blüte der Hersteller war das Standardgeschäft für Acrylglasplatten durch einen hohen Preisdruck und starken Kostenfokus geprägt. Damalige Entscheidungsträger und Betroffene thematisieren die Situation eines preisaggressiven Wettbewerbs bedingt durch die fortschreitenden Commodisierung von PMMA und auslaufende Patente (E).[119] Im Verlauf der Marktreife kam es weiter zum Auseinanderklaffen der Preis-Kosten-Schere durch stark gestiegene Rohölpreise und ein erhöhtes Lohnkostenniveau (E). Weiterhin wirkte sich die Wechselkursentwicklung negativ auf die Geschäftsergebnisse aus (E).[120]

1999-2007 Fusion AuriCo: Strategische Neuausrichtung und Repositionierung. Einen weiteren Meilenstein in der Unternehmensentwicklung der PlastiCo und Wendepunkt zur strategischen Neuausrichtung des PMMA-Geschäfts bezeichnete ein erneuter Führungswechsel, durch Verschmelzung der MineriCo mit dem AuriCo-Konzern (I). Dabei fand eine zunehmende Zentralisierung statt und das Traditionsgeschäft mit PMMA wurde in die Konzernstruktur integriert. In Form zweier weiterer Guss-Produktionsstandorte[121] wurde

[113] In Bezug auf Veränderung in individuellen und kollektiven Wissensstrukturen politischer Entscheidungsträger zu einem veränderten strategischen Verhalten sei an dieser Stelle der Bezug zur Lernperspektive hergestellt.

[114] Weltweit führende Topmanagement-Beratung, in über 50 Ländern tätig (vgl. zur Unternehmensdarstellung ausführlich McKinsey, 2014).

[115] Vgl. #6-19, #99ff., #191ff.,#146, #338ff, D-ii-02, D-ii-03, D-ii-16:9, GmbH, 2007: 96f..

[116] Vgl. #55, #100, zur Ergebnisentwicklung des Sanitärgeschäfts Online-Appendix 6.

[117] Vgl. #52, #101, #190, D-ii-01.

[118] Vgl.#12ff., #19, #110, #192, #267, #354ff., #362, #442, D-ii-02.

[119] Vgl. #43, #145, #220, #230, #245, #226, #292, #295, #297. #300, #365ff., #465ff., D-ii-15:3ff..

[120] Vgl. #46, #155ff., #162ff..

[121] Die hinzugekommenen Standorte in Österreich und Portugal wurden dabei hauptsächlich mit der Produktion von Sanitär-Acrylguss betrieben. Die

eine neue Organisationsstruktur der BU PMMA geschaffen, es änderten sich Zuständigkeiten und Entscheidungsstrukturen. Die Guss- und Extrusionsproduktion wurde fortan bereichsübergreifend geführt, wobei den Aktivitäten aus dem Bereich Qualität, F&E, Marketing und Controlling Zentraleinheiten übertragen wurden. Als Komponente strategischer Prozesse wurde eine jährliche Strategieplanung des BU-Managements und Mittelfristplanung für die drei Folgejahre in Abstimmung mit dem Vorstand durchgeführt.[122] Der Bereichsleiter beschreibt dabei wesentliche Auswirkungen für die Unternehmenspolitik:[123]

"Mit AuriCo kam der Shareholder-Ansatz." (#206)

2000 „MORE". Um neue Wachstumspotenziale zu erschließen und den Konzernzielen zu genügen, wurde durch die neue Unternehmensleitung der AuriCo ein Beratungsprojekt im PMMA-Bereich, unterstützt durch Deloitte Consulting,[124] initiiert (A). Im Rahmen der Entwicklung einer grundlegenden Konzeption für die langfristige Entwicklung des Geschäfts mit PMMA wurde eine weitere Internationalisierung[125] sowie Innovation und Differenzierung angestrebt. Zusammenhängend erfolgte eine zunehmende Ausrichtung am Markt, u. a. unter Umstellung einer zuvor produktorientierten Organisation auf eine Marktsegmentstruktur und Förderung der Innovationsfähigkeit durch Bildung neuer Abteilungen (Business Development).[126]

Spezialitätenstrategie. Das „Standardplattengeschäft" mit PMMA (insbesondere in Kernmärkten des Gussbereichs) der AuriCo war durch die zunehmend angespannte Wettbewerbssituation bei Markteintritt neuer Konkurrenten aus Asien und damit einhergehend weiter sinkenden Margen beeinflusst. Der Spezialchemiekonzern war aufgrund hoher Personal- und Infrastrukturkosten gegenüber spezialisierter Kunststoffproduzenten im „Standard-Geschäft" langfristig nicht wettbewerbsfähig.[127] Die strategische Neuausrichtung mit dem Fokus auf Spezialprodukte („Spezialitäten") bezeichnet das Aufkommen der „Spezialitätenstrategie". Als Form einer Repositionierung des PMMA-Geschäfts wurde eine Transformation der Kernprozesse und Ressourcen zur Besetzung lukrativer Marktnischen (gekennzeichnet durch deutlich höhere Wachstumserwartungen und Margen) bei kontinuierlicher Rücknahme der Marktposition in niedrigpreisigen Segmenten für Standardprodukte (Sanitär) oder deren kostengünstigen Produktion im Ausland angestrebt.

Zur Neuaufstellung als Marktführer im „Spezialitäten-Geschäft" wurden die Aktivitäten in den Bereichen Marketing und Vertrieb ausgeweitet (A), wobei u. a. eine große Marketingkampagne (2003) durchgeführt, vermehrt Messestände betrieben und der PMMA Online-Shop (2006) zum Ausbau des Direktvertriebs gegründet wurde. Es sollten das Händlergeschäft als Hauptabsatzweg angekurbelt und Markinformationen über den Händler hinweg, direkt vom Kunden generiert werden. Außerdem wurden intensive Aktivitäten der Entwicklung des extrudierten Produktportfolios zur Erschließung attraktiver, neuer Marktchancen und der Etablierung als Innovationsführer im Acrylglasmarkt betrieben (A). Im Zuge einer zunehmenden Internationalisierung wurden JV zur Fertigung von PMMA in Russland (A) und China (A) geschlossen. Daneben wurde stetiges Kostenmanagement zur Erreichung anspruchsvoller Rentabilitäts-Ziele des AuriCo-Konzerns betrieben.[128]

2000-2006 Marktniedergang Sanitärmarkt und Versuche zur Revitalisierung. Die zunehmenden Überkapazitäten der Plattenproduzenten und Eintritt neuer Marktteilnehmer aus Asien führten insbesondere im Sanitär-Marktsegment zu einem deutlichem Preisverfall. Weiterhin belasteten neu entwickelte Substitutions-Werkstoffe[129] den Sanitärmarkt (E). Das preisgleiche Angebot endproduktfertiger Komplettsysteme von Acryldusch- und Badewannen chinesischer Produzenten ab dem Jahr 2002 (E) führte letzten Endes zu einem „Verschwinden der Branche" (#107) für traditionell 3-stufig operierende Badewannenhersteller unter Austritt von Marktteilnehmern über die gesamte Lieferkette:[130]

> *"Man war abhängig von wenigen Kunden und es wurden immer weniger." (#465 Betriebsleiter Extrusion)*

Die AuriCo nahm ihre Wettbewerbsposition im Sanitärmarkt fortwährend zurück.[131] Entsprechende Mengenrückgänge werden in der dargestellten Absatzentwicklung ab 2001 ersichtlich. Daneben wurden Initiativen zur Wiederbelebung des Marktes durch gezielte Produktweiter- bzw. Neuentwicklung und entsprechende Marketing-Aktivitäten betrieben (A):

> *"...Wir haben uns mit Free Flow einen Wettbewerbsvorteil versprochen und wir hatten dann auch XT. Das schlechteste Produkt sollte raus. Es gab Vertriebsleiter, die Chancen gesehen haben." (Bereichscontroller #242)*

zusätzlichen Verkaufsmengen sind ab dem Jahr 1999 in der dargestellten Geschäftsentwicklung zu berücksichtigen. Der resultierende Mengenanstieg vom Jahr 1998 zum Jahr 1999 ist überwiegend auf die neu hinzugekommenen Produktionskapazitäten zurückzuführen.

[122] Vgl. #21, #153, #227, #344f., #405ff., D-ii-16:34f., D-ii-15 bis D-ii-21.

[123] Als Gesamtheit von Zielen, Strategien sowie Grundsätzen einer Unternehmung verwendet werden (vgl. Bamberger und Wrona, 2012a:25.

[124] Eine der weltweit größten Gesellschaft in der Wirtschaftsprüfungs- und Beratungsbranche (vgl. zur Unternehmensdarstellung Deloitte, 2014).

[125] Die Internationalisierung (sstrategie) ist hier in prozessbezogener Sicht, im Sinne einer internationalen Entwicklung des Unternehmens, zu verstehen (vgl. zu begrifflichen Grundlagen der Internationalisierung von Unternehmen z. B. Bamberger und Wrona, 2012a:160ff.).

[126] Vgl. #22, #84, #134ff., #202, D-ii-16:9.

[127] Infolge zusätzlicher Kostenpositionen durch gesetzlich vorgeschriebene Infrastrukturen für Chemieunternehmen (Werkfeuerwehr, Werksarzt, etc.), hoher Chemietarife sowie Lohnkostendifferenzen ggü. den Kunststoffherstellern aus Asien (vgl. #63, #132, #175, #248, D-ii-16-20).

[128] #84f., #125, #134, #250, #407, #471, D-ii-16., D-ii-17, D-ii-21.

[129] In Form kostengünstiger weiterzuverarbeitende Coextrudate als Fremdverbunde aus dem Trägermaterial ABS mit dünner PMMA-Oberfläche (vgl. Sitka und Kappacher, 2003).

[130] Vgl. #30, #63, #107, #159, #164, #295, #300, #468ff., D-ee-03, D-ii-15:4, D-ii-16:24.

[131] Vgl. zur Marktanteilsentwicklung auch Online-Appendix 7.

Neben der Weiterentwicklung von Sanitärguss-Produktreihen (u. a. „Free Flow“[132]) zur Produkt-Differenzierung bzw. als spezielle Nischenanwendung wurde bei zunehmendem Ausbau der Extrusions-Technologie in Wachstumsgeschäften ein Versuch der Überführung in ein extrudiertes Produktportfolio unternommen. Mit der Neueinführung eines extrudierten Sanitär-Typs („XT“[133]) wurde Anfang des Jahres 2003 eine Neupositionierung im Sanitärmarkt angestrebt.[134] Es wurden spezifische Produktvorteile, v. a. die reduzierten Herstellungskosten durch das alternative Extrusions-Verfahren und verbesserte Qualitätsmerkmale gegenüber dem marktetablierten Sanitär-Gussacryl, gesehen.[135] Im Geschäfts-Management wurde geplant, das defizitäre Guss-Produkt durch den neuentwickelten extrudierte Sanitär-Typ in verschiedenen Anwendungen langfristig zu ersetzten (vgl. Abbildung 12).[136]

In einem Memorandum von 2003 äußert sich der damalige Leiter des Extrusionsbetriebs zu den Aktivitäten im neu als „Bad und Wellness“ definierten Marktsegment und Erfolgsaussichten der versuchten Produktüberführung:

> *”Für den Markteinstieg mit XT Sanitär wurden alle Register hinsichtlich Bemusterung, technischer Beratung und Dokumentation/Promotion gezogen. Das Fehlen von starken Produktvorteilen hat dennoch dazu geführt, dass der Einstieg mit XT Sanitär in den Sanitärmarkt die Erwartungen deutlich verfehlt hat. Nach Bemusterungen in 2003 wurden in 1-11/ 2004 lediglich 27 to abgesetzt, von denen auch wieder ein Teil nur Bemusterungen waren. Kurzfristig ist mit einer Ausweitung des Geschäfts in den Kilotonnen-Maßstab nicht zu rechnen. Daher besteht die Anforderung seitens PG-IP-Leitung die Marketingaktivitäten zur Einführung von XT Sanitär einzustellen und damit PG-IP nicht weiter mit direkten Marketingkosten in der Größenordnung von ca. 100 TEuro zu belasten. Da das Marktsegment Bad und Wellness neben der Produktgruppe XT Sanitär nur noch die restrukturierungsbedürftige Produktgruppe GS Sanitär vertritt, ist eine Fortführung des Marktsegments Bad und Wellness generell zu überdenken.” (D-ii-13)*

Der extrudierte Produkttyp sowie alternative Guss-Varianten konnten bei bestehenden Ersatzprodukten nicht erfolgreich am europäischen Sanitär-Markt platziert werden und das Revitalisierungskonzept scheiterte. Auch stellten sich Aussichten zur Erschließung neuer Absatzmärkte in Asien und Australien hinsichtlich des Preisniveaus nicht als vielversprechend heraus.[137] Die Produktionsmengen wurden weiter reduziert und im Zuge dessen 2004 das Tochterunternehmen PortoCo als Produktionsstandort für Sanitärguss in Portugal verkauft (A).[138]

2006 „GAP“. Zusammen mit Bearing Point[139] wurde 2006 ein weiteres Restrukturierungsprojekt zur nachhaltigen Verbesserung der Profitabilität und Realisierung identifizierter Wachstumspotentiale des Halbzeuge-Geschäfts mit PMMA durchgeführt. Das „Growth and Profitability“-Projekt erfolgte als „zweiter großer Schritt“ (#112) im Strategieprozess zum vollständigen Rückzug aus dem Sanitärgeschäft und der Neuausrichtung des Geschäftsportfolios mit stärkerer Fokussierung auf profitable Spezialitäten-Geschäfte.[140]

Im Sanitär-Markt war das Preisniveau verfallen, verstärkt zunehmende Preiserosionen in weiteren Kernmärkten des Standardgeschäfts waren mittelfristig zu erwarten. Erarbeitete Maßnahmen im Lenkungsausschuss zielten auf die nachhaltige Ausrichtung auf „Spezialitäten“ ab (vgl. zur Wettbewerbssituation im „Standard-Geschäft“ D-ei-01:4, im Vergleich zum „Spezialitäten“-Geschäft D-ei-01:5f). Neben der Optimierung des Geschäftsportfolios wurden Kernprozesse und Ressourcen in Bereichen des Innovation-Managements, Pricing und Customer Managements angepasst und eine effiziente Kosten- und Kapitalstruktur geschaffen. 2006 war „Gießlinie 3“ vollständig abgeschrieben und es erfolgte die Entscheidung zur Umstellung auf einen halbkontinuierlichen Schicht-Betrieb (A).

Der Bereichsleiter beschreibt wesentliche Implikationen:

> *”Dem ist dann Sanitär zum Opfer gefallen, weil es nicht rentabel war.“ (Bereichsleiter #194)”*

Neben der Desinvestition des Sanitärgeschäfts, unter Einstellung der gesamten Sanitär-Produktpalette (gegossen und extrudiert), wurden weitere, mengenmäßig unbedeutendere Standardproduktreihen eingestellt und das Kunden- und wechselseitig Länderportfolio entsprechend angepasst.[141]

Der Mengenrückgang geht aus der Absatzentwicklung im Geschäftsjahr 2007 hervor. Die Reduzierung der abgesetzten Sanitärplatten auf die Menge „Null“ im laufenden Geschäftsjahr 2008 bezeichnet hierbei den vollständigen Marktaustritt (siehe D-ii-25: „Ausstieg vollzogen“).

[132] PMMA GS SW FREE FLOW als Sanitärguss-Variante mit verbesserten Fließeigenschaften für leichtere Weiterverarbeitung beim Sanitärwarenhersteller sowie PMMA GS SW SAFE mit rutschhemmender Oberflächenstruktur für Nischenanwendung in Krankenhäusern (vgl. #38ff., #88, #115, #209ff., #308, #390; siehe zur Produktbroschüre D-ie-02).

[133] PMMA XT SW RESIST als coextrudierter Sanitär-Produkttyp mit erhöhtem Schlagzähanteil (siehe zu Produktbroschüren D-ie-03, D-ie-04, D-ie-05, D-ie-06).

[134] Das ursprüngliche Sanitär-Segment wurde im Zuge dessen neu definiert und fortan unter der Bezeichnung „Bad und Wellness“ geführt.

[135] Vgl. D-ii-13

[136] Es sei darauf hingewiesen, dass die Produktneueinführung im Rahmen der Auszeichnung als geprüfter „Sanitär-Werkstoff“ durch Produkttests und Spezifizierungsprozesse nach den diversen Produktnormen (vgl. 3.2.3) erschwert war (vgl. #88).

[137] Vgl. #38ff., #88ff., #115, #329.

[138] Vgl. #49f., #104, #470.

[139] Eine in Europa führende Management- und Technologieberatung (vgl. zur Unternehmensdarstellung (BearingPoint, 2014)

[140] Vgl. #51, #127, #233, D-ei-01.

[141] Vgl. #113f., #127, #193, D-ei-01:12ff., D-ii-25, D-ii-26.

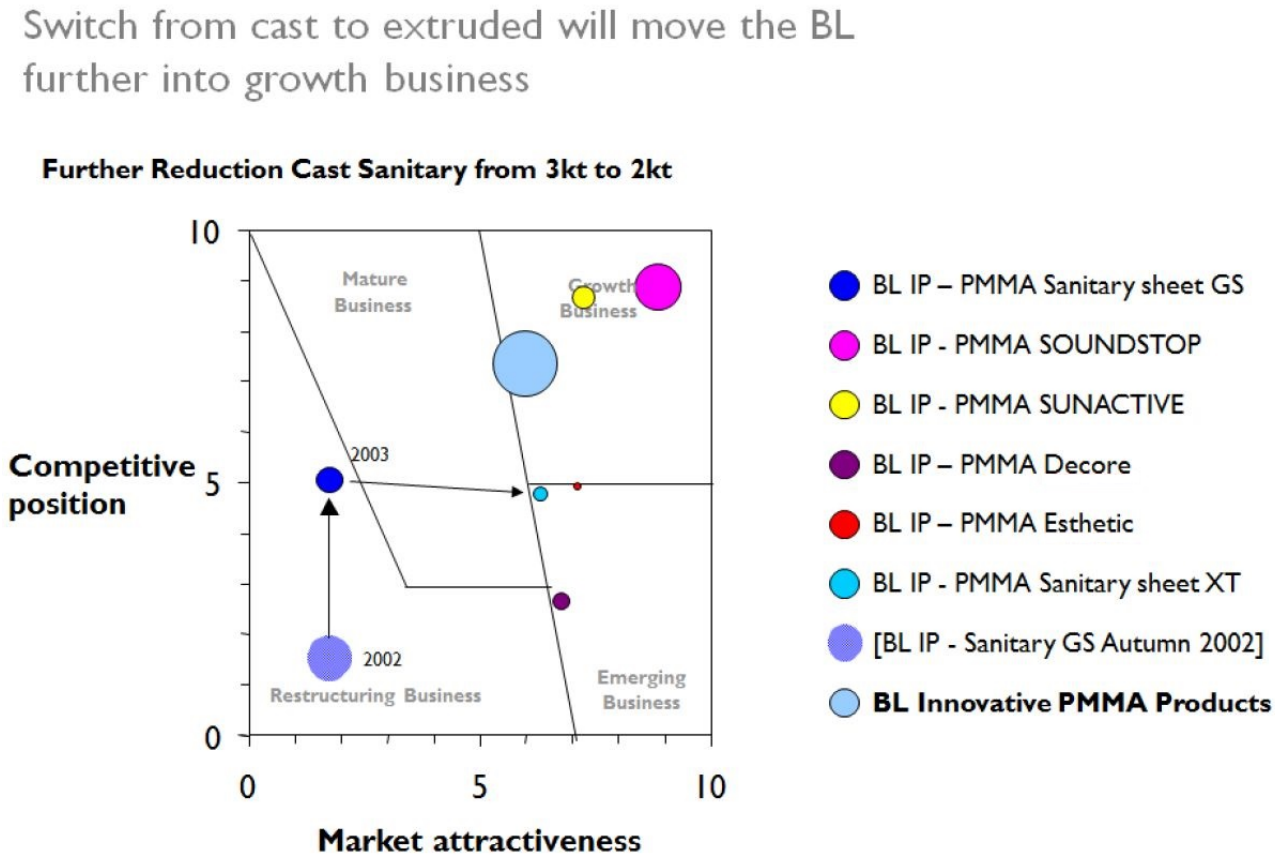

Abbildung 12: Entwicklungsstrategie des Produktportfolios BL Innovation Products (IP) 2003. Bei langfristiger Desinvestition des gegossenen Sanitär-Produkttyps (Sanitary Sheet GS) wird mit dem extrudierten Typ (Sanitary Sheet XT) eine Wachstumsstrategie und Neupositionierung als „Emerging Business" im Produktportfolio verfolgt. (Quelle: Vorstandspräsentation Strategieplanung BU PMMA 2003, D-ii-16:25)

Zwischenfazit zum Prozessverlauf des Rückzugs aus dem Sanitärgeschäft

Die bisherigen Ergebnisse zur Verlaufsform des Rückzugsprozesses aus dem Sanitärgeschäft lassen sich wie folgt zusammenfassen:

Im Rahmen der dynamischen Entwicklung von Wettbewerbern und Substitutionsprodukten im langfristig schrumpfenden Sanitärmarkt ergab sich, dass das Geschäft spätestens seit Anfang der 90er Jahre für PlastiCo unrentabel wurde. Nach dem Aufgang des mittelständischen Familienunternehmens der PlastiCo im Chemie-Großkonzern und als Reaktion auf die Ertragskrise erfolgte im Zuge einer umfassenden Unternehmensrestrukturierung im „TOPFIT"-Projekt 1996 eine Refokussierung des Halbzeuge-Geschäfts mit PMMA, wobei ein langfristiger Rückzug aus dem unprofitablen Direktkunden-Geschäft forciert und zuerst ein selektiver Rückzug aus dem Sanitär-Marktsegment vollzogen wurde. Im Anschluss daran wurden Aktivitäten im Rahmen einer Deckungsbeitragsbetrachtung hingegen wieder gesteigert und nach erneutem Führungswechsels durch Fusionierung und unter zunehmend veränderten Markt- und Wettbewerbsbedingungen in Kernmärkten von Acrylglas ein Versuch zur Umrüstung auf einen neuen Produkt-Typ und Neupositionierung im Sanitärmarkt unternommen. Durch das Scheitern der versuchten Produkt-Überführung wurde die Wettbewerbsposition endgültig zurückgenommen und das Marktsegment im weiteren Entwicklungsverlauf, innerhalb eines zweiten Restrukturierungsprojekts „GAP", im Geschäftsjahr 2008 letztendlich vollständig desinvestiert. Der Rückzug wurde damit langfristig, über einen Zeitraum von 13 Jahren und dem zwischenzeitlichen Versuch zur Revitalisierung als Teilepisode, vollzogen. Auf die angesprochenen „Schritte" zurückkommend, lässt sich der Rückzugsprozess somit vereinfacht als Dreierfolge aus einem teilweisen Rückzug, durch Beschränkung auf bestimmte Kunden bzw. Produktgruppen, zwischenzeitlich wieder gesteigerten Aktivitäten und versuchter „lebensverlängernder Maßnahmen" und dem endgültigen Marktaustritt bei Desinvestition der gesamten Sanitär-Produktpalette beschreiben (siehe Abbildung 13).

3.3.3. Organisationale Trägheit und strategische Rigidität: Ausprägungsformen und Ursachen

Im Rahmen der Beschreibung des Prozessverlaufs des organisationalen und strategischen Wandels des Halbzeuge-Geschäfts mit PMMA konnte gezeigt werden, dass sich weite Teile zentraler Markt- und Wettbewerbsbedingungen auf

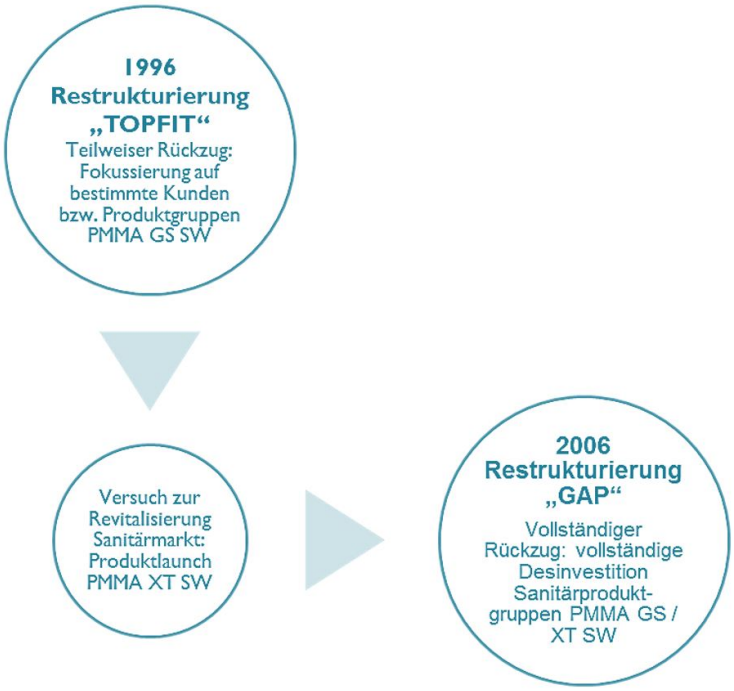

Abbildung 13: Prozessverlauf zum vollständigen Rückzug aus dem Sanitärgeschäft (Quelle: eigene Darstellung)

Acrylgussmärkten grundlegend verändert haben. Anpassungsprozesse zur Realisierung eines beabsichtigten Marktrückzugs hinsichtlich einer defizitären Geschäftslage verliefen hingegen zögerlich:

> *”Es war ein Rückzug auf Raten.” (Director Strategy #60)*

Hierauf Bezug nehmend wurde in den Interviews vermehrt ein „risikoaverses“ und „inkonsequentes“ Entscheidungsverhalten während des Rückzugsprozesses kritisiert, wie das folgende Statement des damaligen Produktionskoordinators verdeutlicht:

> *”Aus heutiger Sicht, vielleicht waren wir zu hasenfüßig, nicht konsequent genug.” (Produktionskoordinator #33)*

Den angesprochenen neuralgischen Punkten soll im Folgenden weiter nachgegangen werden und dazu der Entwicklungsverlauf zum endgültigen Marktrückzug hinsichtlich seines spezifischen Zustandekommens aus verschiedenen Dimensionen auf Ausprägungen und Ursachen von Trägheit untersucht werden.

Historizität. Als Kunststoffhersteller und Vollanbieter von PMMA-Halbzeugen verfolgte PlastiCo traditionell die Strategie der Differenzierung bei Gesamtmarkt-Abdeckung. Das Unternehmen operierte seit Ende der 80er Jahre in einem reifen Markt für Acrylgussplatten, unter dem Druck steigender Überkapazitäten der Plattenproduzenten im Standardgeschäft, wobei eine Wahl zur langfristigen Kapazitätserweiterung durch eine größere Anlageninvestition getroffen wurde.

Interviewteilnehmer, die zu dieser Zeit bereits im Unternehmen tätig waren, heben die historische Vorprägung von Entscheidungsprozessen zum Produktionsausbau und einer speziellen Vertiefung der Gusstechnologie vor dem Hintergrund der vergangenen Markterfolge von PMMA und dabei strategischen Identität als Erfinder und Markführer hervor.[142] In den Interviews wurde eine mangelnde Markt- und Wettbewerbsbeobachtung[143] unterstellt. Ein ehemaliger Betriebsleiter wies dabei auf den starken Einfluss einer primär produktionstechnisch orientierten Geschäftsführung des Familienunternehmens[144] hin:

> *”Wir hatten einen technischen Geschäftsführer, der die neusten Maschinen wollte. Letztlich war das wohl auch der Grund, warum sich die Familie zurückgezogen hat, die hatten keine Vorstellung davon, was am Markt passiert (#454f.)... Wir kaufen Rohstoffe ein und verkaufen Platte.” (Betriebsleiter Extrusion #461)*

Weitergehend wurde eine geringe Geschäftstransparenz in Bezug auf die Ertragslage beschrieben, da die Ergebnisse im damals familiengeführten Unternehmen nicht getrennt erfasst wurden. Nach dem Ausscheiden der Gründerfamilie der PlastiCo aus dem Unternehmen und Verkauf der Anteile an die MineriCo, führten konjunkturell bedingte Umsatzrückgänge und die Ertragskrise zur Offenlegung wirtschaftlicher Ineffizienzen und dem Auslösen von Anpassungsprozessen an eine veränderte Marktumwelt. Unter Einführung

[142] Vgl. #176, #213, #382.
[143] Vgl. #366f., #445f..
[144] Vgl. #124, #184, #454ff., #484.

von Managementsystemen zur Verbesserung der Ergebnis- und Kostentransparenz wurden im Zeitraum zwischen 1994 bis 1996 verschiedene Maßnahmen der Restrukturierung umgesetzt. Ertragsschwache Produkte und Direktkunden-Geschäfte wurden aufgegeben, wobei sich zunächst selektiv aus dem verlustreichen Sanitärgeschäft zurückgezogen wurde.

Austrittsbarrieren. Im allgemeinen Rückzugskontext und insbesondere unter dem Blickwinkel des selektiven Rückzuges sind dahingehend wirksame Austrittsbarrieren näher zu beleuchten. Interviewte wiesen auf potentielle Außenwirkungen eines Ausstiegs als Marktbegründer und marktführender Vollanbieter von PMMA-Halbzeugen als Differenzierungsmerkmal hin.[145] Auch wurden die angesprochenen Verbundeffekte zu anderen Konzerneinheiten verdeutlicht und das bestehende Wechselverhältnis der PMMA-Gussproduktion dabei als „Zwang" zur Abnahme des Vorstufenprodukts MMA herausgestellt, wodurch sich Produkteinstellungen im Gussbereich als prinzipiell problematisch erweisen.[146]

Als weiterhin entscheidender Punkt wurde der Kundenzugang als ökonomische Barrieren angesprochen, sonach ein vollständiger Rückzug (zunächst) durch langfristiger Geschäftsbeziehungen und Lieferverträge behindert war.[147]

Besonders vor dem Hintergrund einer beschrieben langen Geschäftstradition (vgl. 3.2.3) kommen außerdem Ausprägungen emotionaler Barrieren zum Tragen, welche sich rationalen und ökonomischen Kalkülen entziehen. Vergangene Erfolge[148] und ein persönlicher Bezug zur Produkt-Anwendung waren in den Gesprächen im Feld vermehrt zentrales Thema:

> *"Ein emotionales Geschäft, sehr viele Leute hingen daran. Wir haben es selbst entwickelt gegen alle anderen Teilnehmer und stellten es selbst her. Das war eine positive Bestätigung." (Bereichsleiter #142)*

Nachvollziehen ließ sich der emotionale Zuspruch innerhalb des Unternehmens auch anhand der affektierten Veröffentlichungen und Image-Broschüren zur Sanitär-Anwendung (vgl. z. B. D-ie-01, D-ie-05).

Wegen der Tradition im Sanitärgeschäft, der Bedeutung für die Reputation und bei bestehenden strategischen Wechselverhältnissen sind hohe Austrittsbarrieren festzustellen, die einen abrupten Marktausstieg verhinderten und die allgemeine Bereitschaft zum Rückzug im weiteren Entwicklungsprozess beschränkten, selbst wenn niedrige bzw. negative Ertragsraten erwirtschaftet werden.

Spezifische Investition. Zeitgleich der Restrukturierungsepisode ging „Gießlinie 3" in Betrieb. Im Rahmen der Produktionsoptimierung und zur Auslastung der Sollkapazitäten der Gießlinie wurden Produkte verlagert und die Sanitärproduktion bei vorhandenen Überkapazitäten als Deckungsbeitragsgeschäft auf „Gießlinie 3" fortgesetzt.[149] Damit kristallisiert sich ein spezifisches Handlungs- und Entscheidungsmuster zur weiteren Aufrechterhaltung des Sanitärgeschäfts und heraus, wie der damalige Bereichscontroller herausstellt:

> *"Bei Überkapazitäten ist ein Reiz da für ein Deckungsbeitragsgeschäft. Sanitär war eine Art Cash Cow, wurde nur nicht konsequent als solche gefahren." (Bereichscontroller #267)*

Auf gewissermaßen „indirekten" Sunk Costs[150] basierend, bildeten sich durch verschiedene Effekte nun kontingente Entwicklungsprozesse heraus, wonach der beabsichtige Rückzug aus entsprechend unprofitablen Produktionen erschwert wurde. In den durchgeführten Interviews thematisierten Experten des Controllings vor allem Kostenremanenzen infolge erhöhter Abschreibungen der Gießlinie, die fortwährend als starke Wandelbarrieren auf die Entscheidungskultur einwirkten:[151]

> *„... Man hatte Angst vor den Remanenzkosten. Ich bleibe auf fünfzig Prozent der Kosten sitzen. Es ist zwar absehbar, dass der Markt klein ist, aber die Anlagen müssen abgeschrieben sein, da fluten wir nochmal den Markt" (#35 Produktionskoordinator)*

Die Ursachen der bestehenden Kostenremanenz waren v. a. aus politischer und sozialer Perspektive zu betrachten, der damalige Produktionskoordinator erläutert weitergehend kausale Zusammenhänge:

> *"Es gab interne Zwänge (#57)... Strategische Überlegungen [zum endgültigen Rückzug aus dem Sanitärgeschäft] gab es, man hat es aber nicht gemacht. Man hätte die Entscheidung treffen können, aber dann hätte man das Problem gehabt, den ganzen Bereich zu schließen, wegen der Auslastung." (#64f. Produktionskoordinator)*

[145] Vgl. #28, #120, #373, #381f..
[146] Vgl. #16, #254, #385f., #410.
[147] Vgl. #46f., #315, #381f..
[148] Vgl. #118f., #240, #321f., #490.

[149] Die spezifische Investition in „Gießlinie 3" und anschließende Entscheidungsprozesse zur Inbetriebnahme und Vollauslastung wurden rückblickend auch in den durchgeführten Experteninterviews als wegweisende Ereignisse für den weiteren Wandelprozess identifiziert. Vor allem der Investitionsumfang und die spezifische Auslegung der Anlage zur Plattenherstellung in drei festgelegten Standardformaten wirkten sich nachhaltig auf den weiteren Entwicklungsprozess aus.

[150] Indirekt insofern, als dass „Gießlinie 3" nicht spezifisch für die Produktion von Sanitär-Acrylguss, sondern „Standardguss-Produkte" allumfassend vorgesehen war. Im Gegenteil bedingte die spezifische Auslegung der Gießlinie in Bezug auf eine eingeschränkte Flexibilität durch festgelegte Plattenformate sogar signifikante Ineffizienzen bei der Produktion von Sanitär-Platten aus PMMA. Kundenabhängig wurden verschiedene Plattenzuschnitte gefertigt, sodass mit der Produktionsverlagerung hohe Verschnitt-Verluste entstanden, die nicht an den Kunden weitergegeben werden konnten und die positive Ergebniswirkung einer gesteigerten Maschinenproduktivität nahezu nivellierten (vgl. zur „Zuschnitt-Problematik" D-ii-04). Allerdings zeichnete sich die gerade die Sanitärproduktion durch hohe Absatzmengen bei Serienfertigung für die vollautomatisierte Herstellung auf „Gießlinie 3" aus.

[151] Vgl. auch #235.

Unternehmensgrundsätze. Es kann unterstellt werden, dass etwaige Betriebs- bzw. Bereichsschließungen grundsätzlich konfliktträchtig sind. Besonders unmittelbar nach den umbruchartigen Veränderungen der Restrukturierungsepisode mit dem „TOPFIT"-Projekt und dabei massiven Personalabbaus waren betriebsbedingte Kündigungen vorweg politisch schwer durchzusetzen. Betroffene verdeutlichten in den Gesprächen den starken Einfluss des Betriebsrates im Konzern[152] sowie außerdem Aspekte sozialer Rücksichtnahme in Form einer traditionell ausgeprägten Loyalität gegenüber den Mitarbeitern und starken gesellschaftlichen Verantwortung als industrieller Arbeitgeber.[153] Der Bereichsleiter wies konkret auf negative Imageeffekte und Auswirkungen auf die Belegschaft bei entsprechend „harten Schritten" (#197) hin. In dieser Hinsicht ist die hohe Bedeutung der Unternehmensgrundsätze[154] in ihrer spezifischen Funktion zum Transport von Werten und als Beschränkungen für strategische Entscheidungen hervorzuheben.[155]

(Selbst-)verstärkende Effekte. In der Phase zwischen 1996 und 1999 konnten „Erfahrungskurveneffekte" (BCG, 1970) bei der Gussproduktion auf „Gießlinie 3" in Form einer kontinuierlichen Effizienzsteigerung realisiert werden.[156] Im Zuge umgesetzter Debottlenecking-Maßnahmen[157] zur Steigerung der Maschinenproduktivität wurden die bestehenden Kapazitäten der Gießlinie erweitert. Die Ausbringungsmengen konnten kontinuierlich gesteigert und dabei mehr Deckungsbeiträge erwirtschaftet werden. Auf Basis der steigenden Ertragskraft unterblieb der Abbau bestehender Überkapazitäten, die Absatzmengen von Sanitärplatten wurden weiterhin gesteigert.

Einen nächsten Meilenstein im Entwicklungsverlauf markierte die Fusion der MineriCo mit der AuriCo 1999. In Form zusätzlicher Produktionskapazitäten für Acrylguss der beiden Tochterunternehmen der AuriCo (AustriCo in Österreich, PortoCo in Portugal) und Hinzukommen neuer Akteure wurde der Bereich PMMA neu organisiert. Es konnten weitere Portfolio- und Produktionsoptimierung umgesetzt und Kostenvorteile in niedrigpreisigen Segmenten für Standardprodukte erreicht werden. Der Bereichsleiter beschreibt diesen Sondereffekt:

> *"Die Fusionen haben es ermöglicht: die Portfoliooptimierung. (#167)... Die Losgrößen steigen pro Standard, das hat geholfen. Man konnte weiter optimieren und restrukturieren." (Bereichsleiter #169)*

Im Zuge dessen wurde 2001 das höchste Betriebsergebnis für den PMMA-Bereich in der Historie erreicht.[158] Unter Bezug auf den verstärkten Verdrängungswettbewerb in Bereichen des Standardgeschäfts und die Erreichung anspruchsvoller AuriCo-Rentabilitätsziele geriet das Verhältnis der Strategie zur Ausrichtung auf Standardprodukte und der Marktsituation allerdings in ein zunehmendes Missverhältnis: Bei hohem Preisdruck und starkem Kostenfokus war die Wettbewerbsfähigkeit des PMMA-Bereichs durch zusätzliche Kostenpositionen als Chemieunternehmen gegenüber Hauptkonkurrenten spezialisierter Kunststoffhersteller erschwert. Das zugrundliegende strategische Handlungsmuster, Kostennachteile in stagnierenden und schrumpfenden Märkten für „Standard-Acrylglasprodukte" durch stetige Effizienzsteigerung kurzfristig auszugleichen, führte dazu, Bestehendes weiter zu optimieren, statt radikal Neues zu erfinden. Mit der Erweiterung der Gusskapazitäten durch Integration neuer Produktionsstandorte entstanden Größen- und Verbundvorteile[159] bei der Herstellung und im Vertrieb[160], zudem wurden bestehende Komplementaritäten[161] zur MMA-Produktion verstärkt. Die verankerten Technologien und Geschäfte wurden weiter vertieft und sich auf dieser Basis ausgerichtet.

Kernkompetenzen. Eine kontinuierliche Prozess- und Portfoliooptimierung sowie der Sondereffekt der Integration zusätzlicher Produktionskapazitäten dienten dem Aufbau kurzfristiger kostenbasierter Wettbewerbsvorteile und damit der Erreichung eines gesteigerten operativen Ergebnisses als Rückbestätigung des eingeschlagenen Weges zur kosteneffizienten Produktion von „Standard-Acrylglasprodukten". Allerdings verringerte das eingespielte Problemlösungsverhalten und danach weitere Vertiefung etablierter Kernkompetenzen die Fähigkeit sich geänderten Markverhältnissen anzupassen und den „richtigen Absprung" (#269) aus mehr oder minder „bewährten", langfristig schrumpfenden Geschäften zu finden und sich strategisch neu zu positionieren.

[152] Vgl. #34, #238, #273, #424, #481, #510.

[153] Vgl. #24, #65, #197, #247, #270, #481.

[154] Als grundlegende Komponente der Unternehmenspolitik im Sinne allgemeiner Maxime oder Verhaltenslinien für das zukünftige Verhalten und die Entwicklung der Unternehmung. Sie drücken u. a. ethische und moralische Wertehaltungen in Bezug auf den Sinn und Zweck der Unternehmung und ihrer Beziehung zur Gesellschaft aus und definieren Grundhaltungen, an denen sich das strategische Verhalten und die Führung der Unternehmung nach innen orientiert (vgl. Bamberger und Wrona, 2012a:140ff.).

[155] Zusammenhängend mit der Sicherung von Produktionsstandorten und vor dem Hintergrund einer Produktion von Platten aus PMMA im österreichischen Gramatneusiedl wurde im Feld auf die Marienthal-Studie aufmerksam gemacht, worin die sozio-psychologischen Wirkungen von Langzeitarbeitslosigkeit durch Schließung der Textilfabrik Marienthal als Hauptarbeitgeber und Motor der örtlichen Wirtschaft Gramatneusiedl aufgezeigt werden (siehe Jahoda et al., 1975).

[156] Durch fortschreitende qualitative Verfahrenstechniken bei zunehmender Standardisierung und unter Einführung von Gruppenarbeit und Rationalisierungsmaßnahmen (vgl. #391ff.).

[157] Verfahrenstechnische Prozessoptimierung zur Erweiterung der Gesamtanlagenkapazität durch die Beseitigung von Engpässen der Teilanlagen (vgl. #391ff.).

[158] Vgl. zur Geschäftsentwicklung die Vorstandspräsentation der Mittelfristplanung 2002 (D-ii-15:2).

[159] Größenvorteile (Economies of Scale) bei Vorhandensein optimierter, größenbedingter Produktions- und Absatzmengen, Verbundvorteile (Economies of Scope) durch effizientere Herstellung und Vertreiben der Produkte im Verbund.

[160] Die Kostenentwicklung für die Sanitärgussproduktion in Online-Appendix 6 stützt diese Aussage.

[161] Als Verbundeffekte innerhalb der Organisation sowie zwischen Organisationen. Bezogen auf die vertikale Integration im System bewirkt das gesteigerte Produktionsvolumen an Gussacryl einen erhöhten MMA-Bedarf der konzerneigenen Rohstoffproduktion. Die einzelnen Systemeinheiten richten sich zunehmend verstärkt komplementär aneinander aus und es kommt zu weiteren Verflechtung der einzelnen Betriebe und neuen Standorte.

"... Wir wollten uns stärker differenzieren. Wir mussten über Standard hinaus anbieten, Spezialitäten machen. Aber das hat man nicht konsequent durchgesetzt. Wie oft haben wir auch gesagt, wir steigen komplett aus mit Guss" (#407ff. Produktionsleiter Guss)

Unternehmenskultur. In Hinblick auf eine generelle Bereitschaft für Veränderungen wurde das sehr prägnante Handlungsmuster einer vergangenheitsorientierten und risikoaversen „Unternehmenskultur" (#422) bei schwerfälliger Anpassung an sich ändernde Rahmenbedingungen und geteilte Denkschemata einer „kurzfristigen Denke" (#255) im Geschäftsbereich beschrieben. Besonders in Bezug auf eine starke Betriebszugehörigkeit und homogene Zusammensetzung des Managements sowie der Mitarbeiterschaft durch vorwiegend bereichsinterne Stellennachbesetzung lässt sich ein direkter Einfluss vermuten.[162] Der damalige Betriebsleiter des Extrusionsbetriebs und heutige Bereichsleiter beschreibt Wirkungen auf die Implementierung der Rückzugsstrategie:

"Man trennt sich langsam davon, um es konsensfähig zu machen." (#488)

Unter dem Blickwinkel einer klassischerweise eingeschränkten, mit der Hierarchiestufe in Unternehmen weiter abnehmenden Wandlungsbereitschaft und analog sinkenden Veränderungskompetenz[163] wurde in einzelnen Interviews das Vorliegen von Krisensituationen als eine Ausgangsvoraussetzung für entsprechend tiefgreifende Veränderungen angesprochen.[164] Überdies wurden die temporäre Bildung von Projektteams und „Ausbruch" aus dem Tagesgeschäft zur strukturierten Erarbeitung klarer Entscheidungen sowie ein hoher Stellenwert von Anreizsystemen für eine anschließend effiziente Umsetzung thematisiert.[165] Externen Beratern wurde eine unterstützende Funktion innerhalb der Moderation des Transformationsprozess zugesprochen. Zudem wurde eine allgemein gesteigerte Verhaltensakzeptanz von erarbeiteten Veränderungsinitiativen bei einer „neutralen" Legitimierung durch Beratungsunternehmen beschrieben.[166]

Versuche zur Revitalisierung. Im Ergebnis von Willensbarrieren zur Aufgabe der großen Absatzmengen und entstehender Lageveränderungen für das Halbzeuge-Geschäft lassen sich konkrete Aktivitäten zur Produkt-Weiterentwicklung und Überführung in ein extrudiertes Produktportfolio als aktive Widerstände politischer Entscheidungsträger der mittleren Managementebene werten.[167]

"Es war ein Versuch mit extrudiertem Produkt günstigere Herstellkosten zu erzielen. Und man hat eine rutschfeste Guss-Variante entwickelt. Es war klar, glücklich werden wir mit dem Produkt nicht. Wir wollten im Geschäft bleiben und haben weiter Produkte in den Markt gedrückt." (Geschäftsbereichsleiter #209ff.)

Das Konzept zur Entwicklung eines extrudierten Sanitär-Typs wurde allgemein kritisch gesehen:

"... es gab Entwicklungen im extrudierten Bereich. Aber der Markt war geprägt von Guss..." (#308 Vertriebsleiter Sanitär)

"... man hat mal fünfzig Tonnen [Sanitär XT] produziert, aber das war nichts und dann hatte es auch noch nicht mal einen Kostenvorteil. Man ist gescheitert. Es war keine strukturierte Vorgehensweise." (#89 Produktionskoordinator)

Change Management. Angesichts der gescheiterten Markteinführung und gegenwärtigen Wettbewerbsverhältnissen wurde Ende 2003 endgültig der „bewußte Rückzug aus niedrigpreisigen Segmenten für Standardprodukte (Sanitär)" (aus dem Vorstandsbericht 2003, vgl. D-ii-10) im Geschäfts-Management geplant. Der damalige Produktionskoordinator beschreibt das Verhalten bei der Umsetzung:

"2002 gab es die fertige Badewanne zum selben Preis wie die Platte. Dann war nach Strategie gefragt, man wollte raus aus Sanitär, aber man hatte keinen Mut. Das ging dann weiter bis 2006 mit GAP" (#30 Produktionskoordinator)

Mit der Konzipierung des Verkaufs der Gussproduktion in Portugal wurde die Wettbewerbsposition im Jahr 2004 bereits sichtbar zurückgenommen, die offene Kommunikation eines klaren strategischen Kurses und zielgerichtete Steuerung zum Marktaustritt erfolgte allerdings erst im Rahmen der umfassenden Veränderungen in 2006:

"Man hat versucht es so lange mitzuziehen, so lange es sich noch tragen ließ (#383)... Eigentlich war erst gegen Ende hin ganz klar, dass wir aussteigen, da wurde dann auch wieder massiv Personal reduziert..." (#389 Betriebsleiter Guss)

"Man hat gespürt, dass kein Schwung mehr drin war. (#305) ... Latent war immer schon, dass wir aussteigen wollen, es ist nur nie konkret gesagt worden. Es ist immer weniger geworden, man wollte kein Geld mehr investieren." (#313 Vertriebsleiter Sanitär)

"Es war eine Brücke aus dem Markt heraus, hundertprozentig gesteuert war das nie. Man hatte das Gefühl, es ist nicht mehr." (#473 Betriebsleiter Extrusion)

Der vollständige Rückzug wurde schließlich mit Initiierung eines zweiten Restrukturierungsprojekts „GAP" umgesetzt, als „Gießlinie 3" nach zehn Jahren vollständig abgeschrieben war:

[162] Vgl. #247, #264, #273, #383, #415ff., #427.

[163] Vgl. hierzu z. B. die Ergebnisse der Studie des Beratungs- und IT-Dienstleistungsunternehmens Capgemini zum „Change Management" in Online-Appendix 8.

[164] Vgl. #197, #238, #342, #476ff..

[165] Vgl. #49ff., #79ff.., #349.

[166] Vgl. #493ff..

[167] Vgl. #40, #347ff..

”GAP erfolgte dann als zweiter großer Schritt. In GAP war dann mittlerweile das Preisniveau verfallen. Gießlinie 3 war abgeschrieben, man hatte keine Fixkosten mehr und es erfolgte ein Schichtwechsel. . . ” (#112ff. Bereichsleiter)

Zwischenfazit zur organisationalen Trägheit und strategischen Rigidität

Zusammenfassend konnten bei der Untersuchung organisationaler und strategischer Wandlungsprozesse anhand des Rückzugs aus dem Marktsegment für Sanitärplatten aus PMMA über den Analysezeitraum von 1986 und 2008 Beharrungstendenzen festgestellt und dafür ursächliche Veränderungsbeschränkungen identifiziert werden.

Organisationale Trägheit äußerte sich in zwei verschiedenen Ausprägungsformen:

Zunächst ergab sich, dass sich die wandelnden Marktverhältnisse in den 80er Jahren und abzeichnende Marksättigungstendenzen in Segmenten für Acrylgussplatten, nicht (bewusst) erkannt wurden. Erst als diese Anfang der 90 Jahre mit dem Herausstellen der Ertragsschwäche des Sanitär-Marktsegments durch Wechsel der Geschäftsführung reflektiert wurden, und im Zuge der Ertragskrise, erfolgten Anpassungsprozesse und ein Strategiewechsel. Die strategische Kontinuität bei einer eingeschränkten Selbstreflexion zur Wahrnehmung veränderter Markt- und Wettbewerbsverhältnisse kann als eine erste Form von Trägheit angesehen werden.

Innerhalb der gesamtunternehmerischen Restrukturierung können Trägheitstendenzen insofern weiter nachgewiesen werden, als dass Veränderungsinitiativen zum Rückzug aus unprofitablen Geschäften nicht vollständig nach Plan umgesetzt bzw. eine schlussendlich „konsequente“ Umsetzung hinausgezögert wurde. Diese zweite Form von Trägheit zeigte sich darin, dass ein zunächst selektiver Rückzug bei Fokussierung auf profitable Geschäfte nur temporär vollzogen und die Wettbewerbsposition unter Deckungsbeitragsbetrachtung langfristig wieder aufgebaut wurde. Es ließ sich eine zeitweise stark ausgeprägte Wandlungsresistenz - aufgrund verschiedener Wandlungsbarrieren und verstärkender Effekte - feststellen, die sich als offene Form des politischen und emotionalen Widerstandes einflussreicher Akteure gegen die Aufgabe des Sanitärgeschäfts in gezielten Versuchen zur Revitalisierung widerspiegelte. Letztendlich wurde langfristig der vollständige Marktrückzug realisiert.

Es lässt sich schlussfolgern, dass die anfängliche, mangelnde Selbstreflexion zum Erkennen einer Änderungsnotwendigkeit mit Durchführung des Restrukturierungsprojekts „TOPFIT“ in eine andere Form von Trägheit überging. Während es bis Mitte der 90er Jahre nicht zum Auslösen von Veränderungsmaßnahmen kam, manifestierten sich Trägheitskräfte im weiteren Entwicklungsverlauf in einer Einschränkung der Fähigkeit und Bereitschaft zum Wandel und führten zu Verzögerungen des Wandlungsvorhabens.

Zusammenfassend bestätigte sich der Verdacht auf organisationale Trägheit bzw. einen zeitweise ausgeprägten strategischen „Lock-In“.[168] Zum Abschluss der Untersuchung sollen im nun folgenden Kapitel zentrale Erkenntnisse über das spezifische Zustandekommen des Wandel-Phänomens zusammengefasst und vorgefundene Trägheits-Ursachen anhand theoretischer Erklärungsansätze diskutiert werden.

3.3.4. Zusammenfassung der Ergebnisse und theoretische Diskussion

In Anlehnung an den zweistufigen Aufbau der Analyse wurde in einem ersten Schritt der Wandlungsprozess bei Realisierung des Rückzuges aus dem Sanitärgeschäft beschrieben. Der Rückzugsprozess ließ sich durch einen sukzessiven Verlauf, anhand drei spezifisch vollzogener Aktivitäten, charakterisieren.

Dabei war die Geschäftsaufgabe als grundlegender Teil des Wandlungsprogramms zur strategischen Neuausrichtung des Halbzeuge-Geschäfts mit PMMA zu sehen: Als stellvertretend für das schrumpfende „Standard-Geschäft“ spiegelte die Desinvestition den Strategiewechsel von einem Hersteller mit Vollsortiment an Acrylglas-Halbzeugen zu einem Anbieter von vorrangig „Spezialitäten-Produkten“ für lukrative Nischenanwendungen wider. Die beiden thematisierten Typen von Rückzugsstrategien, „Refokussierung“ durch Reduzierung der Anzahl von langfristig schrumpfenden „Standard-Geschäften“ und „Repositionierung“ im Zuge der Neuaufstellung des Geschäfts mit „Spezialitäten“ als Kerngeschäft, waren integrativ im Zusammenspiel zu betrachten.

Bei der Realisierung des Wandels ließen sich Formen von Trägheit und des Widerstandes, im Rahmen eines zögerlichen Handlungs- und Entscheidungsmusters zum Rückzug aus dem schrumpfenden Kerngeschäft und zwischenzeitlichen Versuchen zur Revitalisierung, aufzeigen.

An diesen Befund schloss sich in Schritt 2 eine Ursachen-Diagnose an. Dazu wurde an den zentralen Annahmen des Population Ecology-Ansatzes angeknüpft sowie weiter hilfreiche Erklärungsansätze anderer Perspektiven von kognitiv-interpretativen Ansätze, Theorien zu Kernkompetenzen und Lerntrajektorien sowie Arbeiten zur Pfadabhängigkeit zur Betrachtung herangezogen. Es konnten verschiedene Faktoren während der strategischen Geschäftsentwicklung identifiziert sowie wichtige Beziehungen wirkender Wandelbarrieren und die Art und Weise des Entstehungsprozesses herausgestellt werden.

In den nachfolgenden Kapiteln gilt es, die Ergebnisse der empirischen Befunde zusammenzuführen und spezifische Komponenten und Entwicklungen theoretisch nachzuvollziehen. Dies soll über eine Diskussion des Erklärungsbeitrags ausgewählter Perspektiven zu vorgefundenen Wandel-Phänomenen der Fallstudie bewerkstelligt werden.

[168] In Bezug auf organisationale Trägheit als Resultat pfadabhängiger Prozesse (vgl. hierzu Abschnitt 2.2.3). Die Einstellung eines unterstellten Lock-In-Zustandes und strategische Pfadabhängigkeit wird im nachfolgenden Kapitel noch weiter erörtert.

Structural Inertia

Zur Einordnung der Wandlungsprozesse musste weit vor dem Analysezeitraum, auf die Entstehungsgeschichte des Geschäfts mit PMMA der PlastiCo und Begründung spezifischer Anwendungsfelder zurückgegriffen werden, da die Geschichte wesentlich auf den sich anschließenden strategischen Entwicklungsverlauf ausstrahlte.

Die evolutionstheoretisch motivierte Theorie des „Strucural Inertia“ verweist auf die historische Entwicklung von Organisationen und führt Ursachen von Trägheit auf die Formalisierung und Standardisierung von Zielen und Handlungsmustern in Organisationen zurück. Im Fallgeschehen ließ sich eine starke, den „Satus quo“ stützende „Unternehmenskultur“ als System geteilter Wertevorstellungen und Normen feststellen, die eine grundlegende institutionelle Inflexibilität in Wandelprozessen bewirkte. Eine vorherrschend tief verwurzelte Verbundenheit zum Unternehmen und Traditionsgeschäft sowie hohe Standortzugehörigkeit resultierten in einer stärkeren Bindung an den „Status quo“ und prägten Leitlinien der Planungs-, Produktions- und Steuerungsphilosophie des Geschäfts-Managements. Es wurde ein ausgeprägtes gesellschaftliches Verantwortungsbewusstsein zur Erhaltung von Arbeitsstätten und Standortsicherung deutlich, das Rückzugsprozessen zum tiefgreifenden Wandel generell stark entgegenwirkte. Die anfänglich unvollständige Informationsbasis zur Zeit des eigentümergeführten Unternehmens PlastiCo hinsichtlich der Geschäftslage und Marktsituation, Nachwirkungen früher Entscheidungen, wie im Falle der unwiederbringlichen Anlageninvestition als Sunk Costs-Argument sowie Einflüsse durch hohe Marktaustrittsbarrieren ließen sich weiter als konkrete Faktoren von Formen organisationaler Trägheit feststellen. Speziell die Beharrungstendenzen innerhalb der Wandelprozesse nach der Ära PlastiCo und hierbei unternommenen Aktivitäten zum Aufrechterhalten bestehender Geschäfte, wie v. a. versuchte „lebensverlängernde Maßnahmen“, legten außerdem den Schluss auf den Widerstand politisch einflussreicher Akteure in Verbindung einer unzureichenden Machtbasis des Managements zur Durchsetzung entsprechend „harter Schritte“ nahe, womit besonders die politischen Prozesse als Ursache von Trägheit angesprochen sind. In diesem Sinne lässt sich die strategische Persistenz zum langen Festhalten am unprofitablen Sanitärgeschäft auf das Vorliegen struktureller Trägheit zurückführen. Der populationsökologische Ansatz der Evolutionstheorie liefert ein breites Spektrum von Gründen für vorgefundene Wandlungsbarrieren und vielfältige Anknüpfungspunkte, um das zögerliche Entscheidungsverhalten zu begreifen.

Kognitive Verfestigungen

Für ein tiefergehendes Verständnis spezieller Barrieren zum Erkennen eines Wandlungsbedarfs betrachten kognitiv-interpretative Ansätze organisationale Trägheit aus einer akteursgerichteten Perspektive. Kognitive Verfestigungen beschreiben Phänomene selektiver Wahrnehmung und bestehender Denkschemata, die ein Ausblenden relevanter Umweltveränderungen zur Folge haben können und objektiv notwendigen Veränderungen möglicherweise entgegenstehen. Die Analyseergebnisse zeigten, dass es prinzipiell möglich ist, die Missachtung der Marktsättigungstendenzen zur Zeit der Ära PastiCo durch Prozesse selektiver Wahrnehmung zu erklären, spätestens bei der Benennung der Ertragsschwäche des Geschäfts zu Beginn der 90er Jahre lässt sich die Argumentationslinie jedoch nicht weiter fortführen. Besonders vor dem Hintergrund eines vorwiegend aus den eigenen Reihen besetzten Geschäfts-Managements liefert die Betrachtung geteilter Wissensstrukturen allerdings weiterführend eine interessante Sichtweise zur Erfassung des Trägheits-Phänomens, weil geteiltes Wissen über handlungsleitende Prinzipien ein wesentliches Merkmal „formierter“[169] Strategien, bildet. Die fortwährend vorherrschende Logik einer „kurzfristigen Denke“ und das tradierte „zögerliche“ und „inkonsequente“ strategische Denk- und Verhaltensmuster des (Change) Managements läßt sich hierauf zurückführen.

Kernrigiditäten

Aufschlussreiche Erkenntnisse im Fall identifizierter Rigiditäten bietet ferner das Einnehmen der Perspektive des „Ressource-Based-Views“. Erweiternde Ansätze zu Kernkompetenzen und organisationalem Lernen beziehen sich auf kollektive Lernprozesse, in denen Unternehmen einerseits zum Erfolg führende Fähigkeiten ausbilden und zu Kernkompetenzen verdichten, andererseits historisch gewachsene Kernkompetenzen zu Kernrigiditäten werden, weil Unternehmen trotz Anpassungserfordernissen nicht mehr zu (grundlegender) Veränderung in der Lage sind. Durch exploitative Lernprozesse[170], die auf bestehendem Wissen aufbauen und es von dort aus weiterentwickeln, wird ein experimentelles Vorstoßen in unbekannte Bereiche zugunsten der Replikation und Optimierung bestehender Kompetenzen verhindert. Im Fall der PlastiCo lässt sich der Aufbau der Gusstechnik zum Massenabsatz von Platten aus PMMA nachzeichnen, der zu einer Desensibilisierung gegenüber Marktentwicklungen durch neue Substitutionsprodukte bzw. Wettbewerber und wenig aussichtsreiche Versuche zur Revitalisierung des Marktes durch Produkt-Differenzierung führte. Eine klare Fokussierung auf die Kernkompetenz der Gießtechnik und langwährend verfolgte Ausrichtung auf „Standardguss-Produktionen“ zum Absatz in „Standard-Marktsegmenten“ wie dem Sanitär-Segment, ließe sich somit anhand der Ausweitung der Verfahrenskompetenz ablesen.

Pfadabhängigkeit

Die Pfadtheorie liefert weiter einen Mehrwert zur Betrachtung nachgewiesener Rigiditäten durch Erweiterung zentraler Argumente um eine Prozessperspektive. Um da-

[169] Vgl. hierzu besonders die Ausführung Kirschs im Rahmen seiner Analyse der Genese von Unternehmensstrategien und Relativierung der Bedeutung einer expliziten Strategieformulierung für das Zustandekommen von Strategien der Unternehmung (siehe Kirsch, 1997a).

[170] Exploration und Exploitation als Gegensätze in Bezug auf die Verwendung von Wissen: “Exploration includes things captured by terms such as search, variation, risk taking, experimentation, play, flexibility, discovery, innovation. Exloitation includes such things as refinement, choice, production, efficiency, selection, implementation, execution.” (March, 1991).

durch Hinweise auf kausale Zusammenhänge zwischen Entscheidungen im Entwicklungsverlauf zu gewinnen und eine potentielle Pfadabhängigkeit des Falls PlastiCo zu erörtern, sollen die einzelnen Komponenten, und zwar: Historizität, Critical Juncture, Selbstverstärkung sowie der Lock-In, zur Bestimmung zusammengeführt und der Prozess von der Pfadentstehung bis hin zur Pfadabhängigkeit nachvollzogen werden.[171]

Phase 1: Historizität. Die historische Verankerung von Entscheidungen, wie im Fall der PlastiCo die Investitionsentscheidung zur spezifischen Produktionserweiterung, lässt sich als erste zentrale Komponente eines Pfadentstehungsprozesses auffassen. Festzustellen waren historisch geprägte Strukturen, Prozesse und Verhaltensmuster und eingeschränkte Kontingenz[172] während der Entscheidungsprozesse zum Ausbau der Gusskompetenz und Massenabsatz von Platten aus PMMA durch machtvolle Einflussnahme der Eigentümer und Geschäftsführer der PastiCo.

Phase 2: Critical Juncture und Effekte der Selbstverstärkung. Das Sunk Costs-Argument und anschließende Entscheidungsverhalten zur Inbetriebnahme der „Gießlinie 3“ stellen nun dasjenige „kritische Ereignis“[173] dar, wodurch sich eine pfadbedingte Trägheit herausbildete, die sich innerhalb des Zusammenwirkens verschiedener Effekte zunehmend stark verfestigte und dabei zu einer Art temporären Lock-In-Zustand führte. Durch verriegelnde Prozesse und ein dominant wirkendes Handlungs- und Entscheidungsmuster, welches die Akteure selbst beeinflusste, war eine Veränderung der Lock-In-Situation und dabei der „Absprung“ aus dem Sanitärgeschäft erschwert.[174] Die handelnden Akteure nahmen die Beharrungstendenzen v. a. als Kostenremanenz im Sinne beobachtbarer Symptome dieses ineffizienten, verriegelten Entwicklungsverlaufs war.

In den folgenden Jahren nach Inbetriebnahme der „Gießlinie 3“ - zunächst von 1996 bis 1999 - ließen sich selbstverstärkende Mechanismen[175] in der Form von Erfahrungskurveneffekten, Komplementaritätseffekten und entwickelten Denkschemata in Bezug auf die Planungs- und Steuerungsphilosophie zur Bestätigung des eingeschlagenen Pfades nachvollziehen. Die gesteigerte Ertragskraft bildete dabei das Fundament für eine kontinuierliche Positionierung im zunehmend unattraktiven Kerngeschäft.

Phase 3: Lock-In, verstärkende Effekte und stabilisierende Wandelbarrieren. Zu welchem Zeitpunkt der Übergang in einen Lock-In-Zustand erfolgte, kann schwer bestimmt werden. Strategische Pfadprozesse werden spätestens dann bewusst, wenn ein Wandlungsbedarf vorliegt, der Handlungsspielraum allerdings schon so weit eingeschränkt ist, dass ein Wandel nur noch schwer möglich ist. Vor diesem Hintergrund kann ein strategischer Lock-In spätestens auf den Zeitraum nach der Fusionierung mit dem AuriCo-Konzern um die Jahrtausendwende terminiert werden, da sich eine zunehmende Verlagerung des Acrylglasmarktes nach Asien abzeichnete und der Zusammenbruch der traditionellen Sanitärbranche nicht mehr aufzuhalten war. Eine grundlegende Abweichung vom Kerngeschäft mit „Standardguss-Produkten“ wurde dabei durch den Sondereffekt hinzugekommener neuer Gussproduktions-Standorte im Zuge der Fusion erschwert. Die Identität als selbst entwickeltes Geschäft über mehrere Generationen und dabei dominierend homogene Zusammensetzung der Institutionsmitglieder trugen gleichermaßen unterstützend und stabilisierend zur Aufrechterhaltung des „Status Quo“ bei.

Die Tatsache eines geplanten Wechsels auf eine alternative Herstellungs-Technologie mit der versuchten Überführung in ein extrudiertes Produktportfolio, lässt auf die Einsicht schließen, strategischen Wandel herbeiführen zu wollen. Dies erfolgte jedoch mit so starker Verzögerung, dass (asiatische) Wettbewerber dem, bei zunehmender Modell- und Materialflut, bereits zuvorgekommen waren und die strategischen Reaktionen liefen ins Leere.

Pfadbrechung.[176] Die eingetretene Lock-In-Situation, basierend auf Sunk Costs und einhergehend mit spezifischen emotionalen, kognitiven und sozialen Mustern zur Aufrechterhaltung tradierter Geschäfte, konnte langfristig, durch Neuformierung der Geschäftseinheiten und Fokussierung von „Spezialitäten“, „aufgelöst“ werden. Eine zunehmende Rücknahme der Position in „Standard-Geschäften“ und der dabei endgültige Austritt aus dem Sanitärgeschäft wurden in einem spezifischen Wandlungsprojekt bei der besonderen Konstellation, als das zentrale Sunk Costs-Argument nach vollständiger Abschreibung der Anlage abgeschwächt war, umgesetzt. Im Sinne der erfolgreichen Reformbemühungen kann somit von einem De-locking („path breaking“) der pfadbedingten Trägheit gesprochen werden.

[171] Basierend auf der Konzeptualisierung nach Schreyögg et al. (2003) lässt sich organisationaler Pfadabhängigkeit anhand einzelner Komponenten abbilden, die zugleich in chronologischer Abfolge drei spezifische Phasen (Historizität, positive Rückkopplung, Lock-In) im Pfadentstehungsprozess beschreiben. Mit Eintreten des Lock-In liegt pfadbedingte Rigidität vor, die sich in der trägen Gleichförmigkeit des Entscheidungsverlaufs äußert (vgl. Schreyögg et al., 2003).

[172] In Form von Ergebnisoffenheit und Entscheidungspluralität. Kontingenz meint, „dass historische Entwicklungen offen sind und nicht in irgendeinem Sinn ex ante festgelegt.“ Ackermann, 2001:2f..

[173] Das „Critical Juncture“ bezieht sich auf das wegweisende Ereignis bzw. mehrere Ereignisse, deren Zusammenspiel den Übergang der historischen Vorprägung als erste Phase im Pfadentstehungsprozess zur zweiten Phase von spezifischen Selbstverstärkungsmechanismen einleitet, wodurch schließlich die dritte Phase der Pfadabhängigkeit erreicht wird (vgl. Schreyögg et al., 2003).

[174] Vor allem der Investitionsumfang und die spezifische Auslegung der Anlage wirkten sich nachhaltig auf das Folgeverhalten aus. Die „kritische“ Wirkung ist hierbei nur aus einer retrospektiven Betrachtung zu ermitteln - gerade weil zum Zeitpunkt der Entscheidung noch keine bedeutenden Auswirkungen zu erahnen sind, die zu ineffizienten Entwicklungen führen.

[175] Durch die Wirkkraft von Selbstverstärkungsmechanismen wird der Übergang zum Lock-In Zustand erreicht. Hierzu benennt die Theorie der Pfadabhängigkeit unterschiedliche technologisch bedingte sowie institutionell geprägte Formen (vgl. als Überblick z. B. Sydow et al., 2005:7, Koch, 2004:6).

[176] In der aktuellen Pfadforschung finden sich erweiternde Ansätze, die sich mit der Auflösung von Pfaden befassen. Bei der Pfadbrechung wird prinzipiell zwischen einer exogene und endogene Betrachtungsweise differenziert, wobei in dieser Arbeit diejenigen endogenen Prozesse im Fokus stehen, in denen Akteure bewusst und eigenständig agieren, um den vormals entstandenen Pfad zu brechen (Sydow et al., 2009).

Es lässt sich festhalten, dass einzelne Trägheits-Phänomene im Fallgeschehen durch ein breites Spektrum von Forschungsansätzen erklärt werden können. Die betrachteten Theorien des Structural Inertia, kognitiv-interpretativer Ansätze, Kernkompetenzen und Lerntrajektorien sowie Arbeiten zur Pfadabhängigkeit thematisieren unterschiedliche Inhalte, die sich prinzipiell nicht widersprechen, sondern vielfältige Sichtweisen zur Erfassung von Trägheitsursachen bieten.

Allen Theorien ist dabei eine zeitliche Betrachtungsperspektive gemein, wobei klare Faktoren und kausale Wirkzusammenhänge zur Entstehung von Trägheit benannt werden. Die Pfadtheorie hebt sich insofern weiter ab, als dass sie Trägheit aus dem spezifischen Entwicklungsprozess heraus begründet und einzelne Prozessschritte miteinander verbindet. So mag das Sunk-Cost-Argument bspw. hauptursächlich für Trägheit sein, lieferte alleine jedoch keine hinreichende Erklärung für das Prozessverständnis einer entstandenen Pfadabhängigkeit. Eine pfadbedingte Trägheit bildet sich erst durch das anschließende Entscheidungsverhalten heraus, weil bereits erhebliche Investitionen in die Ausgestaltung und Implementierung genau dieser Strategie geflossen sind.[177] Diskutierte (selbst-)verstärkende Effekte bewirkten dabei wesentliche Verzögerungen im Rückzugsprozess. Es kann vermutet werden, dass eine beabsichtigte Desinvestition ohne die Fusionierung und damit neu hinzugekommener Hauptproduktionsstandorte für betreffende Produkte nicht in diesem Ausmaß auf politische Widerstände gestoßen und dabei versuchte Maßnahmen zur Revitalisierung möglicherweise nicht initiiert worden wären. Gerade war es auch der aktive Versuch einer Neupositionierung, welcher, durch eine mangelnde Kommunikation bis zum schlussendlich radikalen Marktaustritt, zu Unsicherheiten bei Betroffenen über die intendierte Geschäftsentwicklungsstrategie führte. Auf der anderen Seite wirkte die Fusion, bei einer veränderten Unternehmenspolitik und unter dem Druck starker Konzernvorgaben- und Erwartungen, als Treiber zur strategischen Neupositionierung und mithin langfristigen Umsetzung des Rückzugs aus schrumpfenden Kernbereichen.

Die vorliegenden Ergebnisse verdeutlichen die eingangs angesprochene Ambivalenz der hier analysierten Trägheitsfaktoren und implizieren eine Betrachtung von Strategie- und Wandelprozess in einem gemeinsamen Entwicklungsprozess.

Dazu haben Bamberger und Wrona (2012b) ein Instrumentarium in Form eines konzeptionellen Bezugsrahmens zu strategischen Prozessen entwickelt, worauf die empirische Untersuchung stützte. Innerhalb dieses theoretischen Rahmengerüsts war eine adäquate Erfassung vielschichtiger strategische Wandelprozesse und umfassende Erhebung komplexer sozialer Phänomene möglich.

[177] „Path dependence, however, is supposed to mean more than the mere existence of timeworn routines, cognitive rigidities, or structural inertia. It is, first of all, a process." (Sydow et al., 2009: 690).

4. Fazit und Implikationen

Die vorliegende Masterarbeit fokussiert auf eine Analyse organisationaler und strategischer Wandel-Phänomene bei der Realisierung von Rückzugsstrategien anhand des Fallbeispiels des Halbzeuge-Geschäfts mit PMMA der ehemaligen PlastiCo in den Jahren von 1986 bis 2008.

Ziel der empirischen Analyse war es, auf Grundlage theoretischer Konzeptionen Prozesse strategischen Wandels zu beschreiben und daran explikativ aufzuzeigen, ob und warum Organisationen dazu neigen, beständig gegen Wandel zu sein und langfristig an verlustreichen Strategien festgehalten wird. Ein Schwerpunkt der Untersuchung lag dabei auf der Analyse der zur Trägheit führenden Ursachen.

Vor diesem Hintergrund wurde zunächst ein theoretisches Vorverständnis zum relevanten Sachverhalt erarbeitet. Problembezogen wurde ein konzeptioneller Bezugsrahmen eingeführt, mit Hilfe dessen die empirische Fallstudienanalyse strukturiert durchgeführt werden konnte. Das für die Fallstudienarbeit relevante und herauszuarbeitende Geschehen wurde mittels qualitativen Erhebungen auf Basis von Experteninterviews sowie vielfältiger Dokumente und Archivdaten erfasst.

Zur Beschreibung vollzogener Wandlungsprozesse stellten sich in der Datenanalyse enge Verbindungen zwischen dem Strategieprozess und organisationalen Wandel heraus. Es ergab sich, dass strategische Entscheidungen in den organisationalen Kontext integriert waren, und der strategische Entwicklungsverlauf ließ sich als komplexer und iterativer Anpassungsprozess im Zusammenwirken externer Umwelteinflüsse sowie Veränderungen der internen Rahmenbedingungen und spezifisch vollzogener Aktivitäten der Führung des Unternehmens bzw. des Geschäfts-Managements darstellen. Anhand des Geschäftswandels konnte dabei gezeigt werden, dass Rückzugsstrategien durchaus eine Berechtigung als eigenständiges strategisches Werkzeug besitzen und als bewusste Handlungsalternative zur „strategischen Erneuerung" und nachhaltigen Wertsteigerung in Erscheinung treten können. Im Rahmen längerfristiger Rückzugsprozesse aus unattraktiv gewordenen Kernmärkten und einer zunehmenden Fokussierung auf profitable Nischensegmente wurde ein neues Kerngeschäft etabliert und strategischer Richtungswechsel vollzogen. Unter dem Blickwinkel von Rückzugsprozessen aus schrumpfenden Märkten und simultanen Rekonfigurierung bestehender Ressourcen zur Erschließung identifizierter Wachstumspotentiale und Neuaufstellung des Geschäfts ergeben sich damit Anknüpfungspunkte an das äußerst vielfältige Forschungsfeld der „dynamic capabilities" bzw. das von Brown und Eisenhardt (1999) strategischen Vorgehen des „Patching".

Innerhalb spezifischer Wandlungsprozesse ließ sich allerdings eine unkonstruktive Dynamik erkennen: Sinngemäß des Zustandekommens eines „Rückzugs auf Raten" spiegelten sich in einem langandauernden Rückzugsprozess zum vollständigen Austritt aus dem schrumpfenden Sanitär-Markt und unter zwischenzeitlichen Versuchen zur Revitalisierung konkrete Ausprägungsformen von Trägheit und Widerstand

wider. Mit Blick auf das weitere Forschungsvorhaben, wurden die empirischen Befunde mithilfe der ausgewählten Ansätze: Structural Inertia, Kernkompetenzen und kognitiven Verfestigungen sowie durch pfadtheoretische Annahmen auf beeinflussende (Trägheits-)Faktoren untersucht.

Durch die Analyse wurde eine starke historische Vorprägung beachteter Wandlungsprozesse deutlich. Eine lange Geschäftstradition und standardisierte Prozesse und Strukturen stark verbundener Geschäfte sowie institutionalisierte Routinen und Normen waren grundsätzlich als endogene Einflussfaktoren für die Herausbildung „struktureller Trägheit“ zu sehen. Der theoretisch diskutierte positive Zusammenhang zwischen Unternehmensalter, -größe und -komplexität und der damit zunehmend eingeschränkten organisationalen Flexibilität, ließ sich empirisch bestätigen.[178] Außerdem schien die starke Unternehmenskultur eines vergangenheitsorientierten und risikoaversen Kulturtyps durch emotionale, soziale und kognitive Wandelbarrieren eine ausschlaggebende Rolle zur Aufrechterhaltung des „Status quo“ gespielt zu haben. „Sunk Costs“ in Form von einer neuen Produktionslinie ließen sich schließlich als Hautursache für die Veränderungsbeschränkung benennen. So ließen sich vorgefundene Wandel-Phänomene durch Ursachen organisationaler Trägheit erklären. Dabei konnten speziell auch Einflüsse durch kognitive und politische Prozesse aufgezeigt werden. . Die Investition in Gießlinie 3 als „kritisches Ereignis“ und der so eingeschränkte Handlungsspielraum ließ außerdem die Vermutung auf eine entstandene Pfadabhängigkeit zu. Es wurden verschiedene, die Pfadprozesse kennzeichnende (selbst-)verstärkende Effekte ausfindig gemacht, die zu strategischen Rigiditäten im Entwicklungsverlauf beisteuerten. Ein organisationaler Sondereffekt durch Unternehmenszusammenschlüsse wirkte hierbei besonders verstärkend. Im Zusammenwirken der Wandelbarrieren konnte eine Art Lock-In-Situation zur temporären „Verriegelung“ der Rückzugsentscheidung aus schrumpfenden Kerngeschäften ermittelt werden, die unter spezifischen Bedingungskonstellationen langfristig „aufgelöst“ und der strategische Richtungswechsel damit vollständig vollzogen wurde.

Innerhalb des Prozessverlaufs ließen sich drei Zusammenhänge erkennen: Einerseits wurde die Emergenz von Strategiewechsel und Führungswechseln in der Unternehmensspitze - als in der unternehmerischen Praxis durchaus üblich - und besonders der Effekt eines Wechsel von einem über mehrere Generationen familiengeführten Unternehmen auf ein familienfremdes Management im Großkonzern deutlich.[179] Daneben war ein positiver Zusammenhang zwischen dem Vorliegen von Krisensituationen und dem Auslösen von Anpassungsprozessen an veränderte Marktmerkmale, in Entsprechung der resignierenden Feststellung „Ohne Krise kein Wandel“ (vgl. Krüger, 2002:88) festzustellen. Die Gestaltung eines strukturellen Rahmens in Form eines Veränderungsprojekts und Bildung einer Projektorganisation stellte sich als grundlegend für eine erfolgreiche und effiziente Umsetzung des Wandlungsvorhabens heraus. Insbesondere in Bezug auf etablierte Denk- und Verhaltensmuster war das Hinzuziehen externer Organisationsentwicklungsberater hierbei für eine kritische Selbstreflexion förderlich und wirkte sich weiter positiv auf die Verhaltensakzeptanz der am Wandlungsprozess beteiligten bzw. von ihm betroffenen Akteure und Organisationseinheiten gegenüber den Zielen und Maßnahmen der Veränderungsinitiativen und damit das Umsetzungsverhalten politischer Entscheidungen aus.

Zusammenfassend wurde anhand der Fallstudie deutlich, dass spezifische Wandel-Phänomene und dabei augenscheinliche Ineffizienzen während der Realisierung von Rückzugsstrategien vor dem Hintergrund wirkender Trägheitskräfte erklärt werden können. In Verbindung mit analysierten Verknüpfungen zwischen strategischen Prozessen und Veränderungen des organisationalen Kontext, wurde die Ambivalenz beeinflussender Faktoren, insbesondere beachteter interner Kontextfaktoren, die einerseits als Treiber des Wandels erscheinen, sich gleichsam aber in auch Trägheitstendenzen äußern oder allgemein als Wandelbarrieren wirken, deutlich. Genau dieses „Dualitätsdilemma“ und die gegenseitige Dynamik von Strategie- und Wandelprozessen stellen den zentralen theoretischen Beitrag der vorliegenden Arbeit dar. Ein übergreifendes Fallverständnis war nur über eine integrative Betrachtung strategischer Entscheidungen im organisationalen Kontext möglich, wobei Rückschlüsse auf den Rückzugsprozess im Fallgeschehen gezogen wurden.

Die vorliegenden Analyseergebnisse implizieren, dass die isolierte Betrachtung von Strategie- und Wandelprozess nicht zielführend für das Verständnis strategischer Prozesse ist und eine Trennung von Strategieentwicklung und -implementierung schwer haltbar, da sich der strategische Managementprozess nicht als strikte Abfolge klar trennbarer sequentieller Prozessschritte, sondern eher als iterativer, sich teilweise überlappender Prozess darstellte, der permanent Wandelprozesse integriert.[180] Insbesondere hinsichtlich der Aktivitäten zur Überwindung von Trägheit und Widerständen waren Prozesse der Generierung von Strategien eng miteinander verwoben. Das traditionelle Modell des Wandels „punctuated equilibrium“ (Tushman und Romanelli, 1985, 1994), das Wandel als kurze Ausbrüche radikaler Veränderungen zwischen sonst langen stabilen Phasen versteht, konnte dabei empirisch nicht eindeutig bestätigt werden. Vielmehr sollte der Strategieprozess eine dynamische Dimension und kontinuierlichen Wandel integrieren, wie auch in der Strategieprozessforschung seit den 90er Jahren verstärkt propagiert wird.[181]

[178] Vgl. hierzu v. a. die Ausführungen von Hannan/Freeman.

[179] Siehe hierzu z. B. die empirische Untersuchung zum Führungswechsel und Strategiewandel bei deutschen Großunternehmen von Hungenberg et al., 2006.

[180] Vgl. hierzu insbesondere die Diskussion von Mintzberg (1990) zur „strategy formation“, die er als Alternative zur plandeterminierten Formulierung und der entsprechenden Abgrenzung der Implementierung anführt. Mintzberg argumentiert dabei, dass das Management weder über erforderliche vollständige Informationen verfügt, noch die Zukunft ausreichend zuverlässig prognostizieren kann und kritisiert die Trennung von Strategieformulierung und Strategieimplementierung (Mintzberg, 1990:186).

[181] Vgl. u. a. Mühlbach, 2003; Pettigrew, 1990; Prahalad und Hamel, 1994

Im Rahmen der kritischen Würdigung der Analyseergebnisse müssen zwei Problemfelder angesprochen werden. Neben der sich aus einer Einzelfallstudie ergebenden prinzipiellen Problematik muss vor allem die Zuverlässigkeit der Datenlage kritisch hinterfragt werden. In der vorliegenden Untersuchung wurde auf Tonaufnahmen durchgeführter Interviews verzichtet, das empirische Material bildeten eigenhändig angefertigten Gesprächsnotizen, worauf sich die Datenauswertung im Wesentlichen stützte. Diese Schwachstelle wurde durch Vorlage der Interview-Transkripte sowie der weiter ausgeführten Interpretationen und Zustimmung durch die Interviewteilnehmer versucht zu kompensieren. Um methodische Mängel weiter auszugleichen wurde eine Triangulation anhand alternativer Datenquellen durchgeführt. Hierbei muss allerdings angemerkt werden, dass vorhandene bzw. ehemals existierende interne Dokumente, welche originäre strategische Planungsprozesse und Einschätzungen aufzeigen, nicht vollständig erhoben werden konnten oder vorlagen. Hinsichtlich der aufgetretenen Wirren der Führungswechsel und Umstellungen der IuK-Systeme konnten einzelne Phasen in der Unternehmensentwicklung nicht vollständig erfasst werden und es ergaben sich Lücken in der Datenlage. Die vorliegende empirische Studie ist als Einzelfallstudie konzipiert und eine Generalisierbarkeit auf Basis der Studienergebnisse nicht möglich. Dennoch lässt die Studie die Einschätzung zu, dass strategischer Wandel durch wirkende Trägheitskräfte prinzipiell erschwert ist und diese als besonderer Problembereich bei der Realisierung von Rückzugsstrategien in Erscheinung treten können. Es wurden verschiedene Faktoren empirisch nachgewiesen, die strategischen Rückzugsprozessen entgegenwirkten und dabei langfristig zu wirtschaftlichen Ineffizienzen führen.

Unmittelbar aus den Limitationen ergeben sich erste praktische Implikationen. Die Möglichkeit zur Sensibilisierung für die Existenz und Emergenz von Trägheit muss im Bewusstsein verankert und negative Wirkaspekte im Rahmen strategischer Wandlungsvorhaben in Erwägung gezogen werden. So wurde im Fallgeschehen der begrenzende Charakter interner Kontextfaktoren deutlich. Das (Change) Management ist dadurch zunächst vor die Herausforderung gestellt, eine allgemeine Veränderungsbereitschaft zu schaffen bzw. Akzeptanz und Unterstützung sowie Promotion geplanter Wandlungsprozesse zu entwickeln. Die Bereitschaft und Fähigkeit zum Wandel sind dabei nicht nur Sache des obersten Managements, sondern müssen fester Bestandteil des Selbstverständnisses aller Führungskräfte und Mitarbeiter sein. In der Literatur besteht weitgehend Einigkeit darüber, dass die Einbeziehung aller Prozessbeteiligter einen entscheidenden Erfolgsfaktor für die Implementierung von Wandel darstellt.[182] Im Rahmen von Maßnahmen zur Veränderung von Komponenten individueller und kollektiver Wissensstrukturen kommt dabei der Kommunikation eine zentrale Bedeutung zu. Zusätzlich können Maßnahmen zu Aufbau und Ausübung von Macht und Anreizsysteme nötig sein, um zu langfristig sinnvollen Handlungs- und Entscheidungsmustern überzugehen. Wenn organisationale Trägheit Rückzugsstrategien zum Wandel verzögern kann, existieren sicher weitere Wirkbereiche, denen während zukünftiger Forschung weiter nachzugehen wäre. Denkbar sind Einflussquellen von Trägheit innerhalb strategischer Veränderungen auf funktionaler Ebene, wie bspw. Beschaffung, F&E, Logistik oder Vertrieb sowie speziell in Neuerungen in Bereichen der Produkt- und Prozessentwicklung.

Im beschriebenen Fall, wobei sich eine grundlegende institutionelle Flexibilität aus einer Kombination von emotionalen, kognitiven, sozialen und ressourcenbedingten Ursachen herausgebildet hatte, waren spezifische Konstellationen notwendig, um eine allgemeine Veränderungsbereitschaft herzustellen und Wandlungsprozesse wirksam und erfolgreich umzusetzen. Es kann gemutmaßt werden, dass eine zeitlich effizientere bzw. „konsequentere“ Umsetzung, u.U. auch verbunden mit der frühen Aufgabe eines gesamten Produktions-Betriebs, zu einer grundlegend veränderten (wirtschaftlich effizienteren) Geschäftsentwicklung, geführt hätte. Allerdings muss honoriert werden, dass das PMMA-Geschäft dann nicht in der heutigen Position im ChemCo-Konzern bestehen würde. Die Erfahrungen aus der Praxis zeigen, dass es nicht DEN einen Weg gibt, der zu einer „erfolgreichen“ Veränderung führt. Die Kompetenz, Veränderungen sinnhaft einzuleiten und Rahmen zur Implementierung zu gestalten bleibt eine Schlüsselfähigkeit des Managements.

Abschließend bleibt zu diskutieren, ob mit dem PMMA-Geschäft innerhalb der heutigen ChemCo die vorerst letzte Stufe des Produktlebenszyklus und das endgültige Ende einer langen Restrukturierungsphase erreicht ist, wenn Wettbewerber die verbleibenden Nachfragenischen des Spezialitätengeschäfts weiter besetzen. Im Gegensatz dazu bekräftigt der Bereich Möglichkeiten zum „Tournaround“. Aufgrund der anhaltend wandelnden Umwelt- und Marktanforderungen und unter Berücksichtigung der aktuellen Strategie in einem innovationsführenden Unternehmen als Anbieter von Spezialitätenchemie, wird die Weiterentwicklung der organisationalen Veränderungsfähigkeit dabei zum kritischen Wettbewerbsfaktor. Ein wichtiger Schritt in diesem Sinne ist daher sicherlich, die Fessel von Wandelbarrieren zu erkennen und eine grundlegende Flexibilität „einzubauen“, um einen proaktiven Wandel innerhalb angestrebter Ziele und Zukunftsvisionen zu ermöglichen.

und besonders die Ausführungen von Brown und Eisenhardt (1997, 1998, 1999) zur Abbildung der Dynamik des Strategieprozesses.

[182] Vgl. u. a. Kieser et al., 1998; Greif et al., 2005; Krüger, 2002; Doppler und Lauterburg, 2008.

Literatur

Ackermann, R. *Pfadabhängigkeit, Institutionen und Regelreform*. Mohr Siebeck, Tübingen, 2001.

Al-Laham. *Strategieprozesse in deutschen Unternehmungen*. Gabler, Wiesbaden, 1997.

ANSI. American standard soaking tub, air spa and whirlpool specification sheet, 2005. s.1.:ANSI.

Anslinger, P., Klepper, J., und Subramaniam, S. Breaking up is good to do. *McKinsey Quaterly*, 1:16–27, 1999.

Arrow, K. Path dependence and competitive equilibrium. In *History matters: essays on economic growth, technology, demographic change*, pages 23–35. University of Michigan Press, 2003.

Arthur, W. Competing technologies, increasing returns, and lock-in by. *Economic Journal*, 99(394):116–131, 1989.

ASTM. Standard Specification for Poly(Methyl Methacrylate) Acrylic Plastic Sheet, 2002. s.1.:ASTM.

Axelrod, R. *The structure of decision: The cognitive maps of political elites*. Princeton University Press, 1976.

Bamberger, I. und Delic, J. Rückzugsstrategien : Motive, Formen, Einflussgrößen. *WiSt-Wirtschaftswissenschaftliches Studium*, 39(1):16–22, 2010.

Bamberger, I. und Gabele, E. Prozessanalyse. Die Planung tiefgreifender Veränderungen von Organisationen als Gegenstand theoretischer und empirischer Untersuchungen. *Betriebswirtschaftliche Forschung und Praxis*, 29 (5):452–469, 1977.

Bamberger, I. und Wrona, T. *Strategische Unternehmensführung. Strategien, Systeme, Methoden, Prozesse*. Vahlen, München, 2 edition, 2012a.

Bamberger, I. und Wrona, T. *Strategische Unternehmensberatung. Konzeptionen, Prozesse, Methoden*. Springer Gabler, Wiesbaden, 6 edition, 2012b.

Barney, J. Firm Resources and Sustained Competitive Advantage. *Journal of Management*, 17(1):99–120, 1991.

BCG. The Product Portfolio. Boston: Boston Consulting Group, 1970.

Bea, F. und Haas, J. *Strategisches Management*. Lucius & Lucius, Stuttgart, 5 edition, 2009.

BearingPoint. BearingPoint Über uns, 2014. URL http://www.bearingpoint.com/de-de/7633-493/ueber-uns/? Zugriff am 12.09.2014.

Bettis, R. und Prahalad, C. The dominant logic: Retrospective and extension. *Strategic Management Journal*, 16(1):5–14, 1995.

Beyer, J. *Pfadabhängigkeit. Über institutionelle Kontinuität, anfällige Stabilität und fundamentalen Wandel*. Campus, Frankfurt, 2006.

Blake, R. und Mouton, J. *The Managerial Grid: The Key to Leadership Excellence*. Gulf Publishing, Houston, 1964.

Bortz, J. und Döring, N. *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Springer, Berlin, Heidelberg, 4 edition, 2006.

Bowman, E. und Singh, H. Corporate restructuring: Reconfiguring the firm. *Strategic Management Journal*, 14(1):3–14, 1993.

Brauer, M. What Have We Acquired and What Should We Acquire in Divestiture Research? A Review and Research Agenda. *Journal of Management*, 32(6):751–785, 2006.

Brown, S. und Eisenhardt, K. The Art of Continuous Change: Linking Complexity Theory and Time-Paced Evolution in Relentlessly Shifting Organizations. *Administrative Science Quarterly*, 42(1):1–34, 1997.

Brown, S. und Eisenhardt, K. *Competing on the edge: Strategy as structured chaos*. Harvard Business School Press, Boston Massachusetts, 1998.

Brown, S. und Eisenhardt, K. Patching. Restitching business portfolios in dynamic markets. *Harvard Business Review*, 77(3):72–82, 1999.

Burgelman, R. Strategy as vector and the inertia of coevolutionary lock-in. *Administrative Science Quaterly*, 47(2):325–358, 2002.

Byerly, R., Lamont, B., und Keasler, T. Business portfolio resturcturing, prior diversification posture and investor reactions. *Managerial and Decisioin Economics*, 24(8):535–548, 2003.

Chandler, A. *Strategy and Structure. Chapters in the History of the Industrial Enterprise*. MIT Press, Cambridge, 1962.

ChemCo. Unternehmensdarstellung der ChemCo AG. ChemCo AG, 2014a. URL http://corporate.chemco.com/de/unternehmen/pages/default.aspx. Zugriff am 11.07.2014.

ChemCo. Geschichte der PlastiCo GmbH. ChemCo AG, 2014b. URL http://geschichte.chemco.de/sites/geschichte/de/gesellschaften/roehm/pages/default.aspx. Zugriff am 09.07.2014.

ChemCo. Anwendungsbereiche PMMA. ChemCo AG, 2014c. URL http://www.PMMA.de/product/PMMA/de/anwendungsbereiche/pages/default.aspx. Zugriff am 11.07.2014.

ChemCo. Struktur und Organisation der ChemCo AG. ChemCo AG, 2014d. URL http://corporate.chemco.com/de/investor-relations/unternehmen/struktur-und-organisation/Pages/default.aspx/default.aspx. Zugriff am 11.07.2014.

ChemCo AG. Technische Daten PMMA, 2013. URL http://www.PMMA.de/product/PMMA/de/produkte/technische-daten/pages/default.aspx. Zugriff am 18.08.2014.

Darwin, C. *On the Origin of Species by Means of Natural Selection, Or, The Preservation of Favoured Races in the Struggle for Life*. J. Murray, London, 1859.

David, P. Clio and the economics of QWERTY. *The American Economic Review*, 75(2):332–337, 1985.

Decker, C. und Mellewigt, T. Thirty Years After Michael E. Porter: What So We Know About Business Exits? *Academy of Management*, 21(2):41–55, 2007a.

Decker, C. und Mellewigt, T. The Drivers and Implications of Business Divesture - An Application and Extension of Prior Findings, 2007b. URL http://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2007-054.pdf. Zugriff am 21.05.2014.

Decker, C. und Mellewigt, T. Business Exit and Strategic Change: Sticking to the Knitting or Striking New Path? *British Journal of Management*, 23: 165–178, 2012.

Deloitte. Über Deloitte. Deloitte Touche Tohmatsu Limited, 2014. URL http://www2.deloitte.com/de/de/pages/about-deloitte/articles/about-deloitte.html. Zugriff am 05.09.2014.

Dohm, L. *Die Desinvestition als strategische Handlungsalternative: Eine Studie des Desinvestitionsverhaltens*. Dissertation, Universität Frankfurt am Main, 1988.

Doppler, K. und Lauterburg, C. *Change Management: Den Unternehmenswandel*. Campus, 12 edition, 2008.

Dougherty, D. Managing your core incompetencies for corporate venturing. *Entrepreneurship Theory and Practice*, 19(3):113–135, 1995.

Easton, D. *A Systems Analysis of Political Life*. Wiley, New York, 1965.

Eisenhardt, K. Building Theories from Case Study Research. *The Academy of Management Review*, 14(4):532–550, 1989.

Eisenhardt, K. und Martin, J. Dynamic capabilities: What are they?! *Strategic Management Journal*, 21:1105–1121, 2000.

Eschenbach, R. Controlling. Stuttgart: s.n., 1995.

Fahey, L. und Christensen, H. K. Evaluating the research on strategy content. *Journal of Management*, 12(2):167–183, 1986.

Flick, U. Triangulation. In *Handbuch Qualitative Sozialforschung*, pages 432–434. Beltz, Psychologie Verlags Union, Weinheim, 1995.

Flick, U. *Qualitative Sozialforschung: Eine Einführung*. Rowohl, Reinbek, 3 edition, 2005.

Gälweiler, A. *Unternehmensplanung. Grundlagen und Praxis*. Campus, Frankfurt am Main, 1974.

Gälweiler, A. *Strategische Unternehmensführung*. Campus, Frankfurt am Main, 3 edition, 2005.

Geiser, J. *Wachstumshemmnisse mittelständischer Industriebetriebe*. Otto Schwartz & Co, Göttingen, 1983.

Glaser, B. *Theoretical Sensitivity*. The Sociology Press, 1978.

Glaser, B. und Strauss, A. *Awareness of Dying*. Aldine Pub. Co, Chicago, 1965.

Glaser, B. und Strauss, A. *The discovery of grounded theory: strategies for qualitative research*. Aldine Pub. Co, Chicago, 1967.

GmbH, P. 100 Jahre Zukunft. Die PlastiCo GmbH von 1907 bis 2007. Essen: PlastiCo GmbH, 2007.

Gomez, G. und Zimmermann, T. *Unternehmensorganisation. Profile, Dynamik, Methodik*. Campus, Frankfurt am Main, 1992.

Gomez, P. und Müller-Stewens, G. Corporate Transformation. Zum Management fundamentalen Wandels großer Unternehmen. In *Unternehmerischer Wandel : Konzepte zur organisatorischen Erneuerung*, pages 135–198. Gabler, Wiesbaden, 1994.

Greenwood, R. und Hinings, C. Organizational Design Types, Tracks and the Dynamics of Strategic Change. *Organization Studies*, 9(3):293–316, 1988.

Greif, S., Runfe, B., und Seeberg, I. *Erfolge und Misserfolge beim Change Management*. Hogrefe, Göttingen et al, 2005.

Habel, S. *Strategische Unternehmensführung im Lichte der empirischen Forschung. Bestandsaufnahme und kritische Würdigung eines komplexen Forschungsfeldes*. Kirsch, Herrsching, 1992.

Halme, M. Corporate Environmental Paradigms in Shift: Learning During the Course of Action at UPM–Kymmene. *Journal of Management Studies*, 39(8):1087–1109, 2002.

Hannan, M. und Freeman, J. The Population Ecology of Organizations. *American Journal of Sociology*, 82(5):929–964, 1977.

Hannan, M. und Freeman, J. Structural Inertia and Organizational Change. *American Social Review*, 49(2):149–164, 1984.

Hannan, M. und Freeman, J. *Organizational Ecology*. Harvard University Press, 1989.

Harrigan, K. R. Exit decisions in mature industries. *Academy of Management Journal*, 25(4):707–732, 1982.

Harrigan, K. *Strategies for declinging businesses*. Lexington Books, Toronto, 1980.

Harrigan, K. Deterrents to Divesture. *Academy of Management Journal*, 24 (2):306–323, 1981.

Harrigan, K. und Porter, M. End-Game Strategies for Declining Industries. *Academy of Management*, 61(4):111–120, 1983.

Harrison, F. Perspectives on Strategic Management. *International Journal of Management*, 7:437–445, 1990.

Herbert, T. und Deresky, H. Generic strategies. An empirical investigation of typology vadility and strategy content. *Strategic Management Journal*, 8 (2):135–147, 1987.

Hilker, J. *Marketingimplementierung: Grundlagen und Umsetzung am Beispiel ostdeutscher Unternehmen*. Deutscher Universitätsverlag, Wiesbaden, 1993.

Hoskisson, R. und Johnson, R. Corporate refocusing and strategic change: The effect on diversification strategy and R&D intensity. *Strategic Management Journal*, 13(8):625–634, 1992.

Hoskisson, R., Johnson, R., und Moesel, D. Corporate divestiture intensity in restructuring firms: Effects of governance, strategy, and performance. *Academy of Management Journal*, 37(5):1207–1251, 1994.

Hoskisson, R., Johnson, R., Tihanyi, L., und White, R. Diversified Business Groups and Corporate Reffocusing in Emerging Economies. *Journal of Management*, 31(6):941–965, 2005.

Huff, A. und Reger, R. A Review of Strategic Process Research. *Journal of Management*, 13(2):211–236, 1987.

Huff, J., Huff, A., und Thomas, H. Strategic renewal and the interaction of cumulative stress and inertia. *Strategic Management Journal*, 13:55–75, 1992.

Hungenberg, H., Wulf, T., und Stengl, K. Führungswechsel und Strategiewandel: Eine empirische Untersuchung bei deutschen Großunternehmen. 60(2):141–155, 2006.

Institut für Mittelstandsforschung. KMU-Definition des IfM Bonn, 2014. URL http://www.ifm-bonn.org/mittelstandsdefinition/definition-kmu-des-ifm-bonn/. Zugriff am 21.08.2014.

Jahoda, M., Lazarsfeld, P., und Zeisel, H. *Die Arbeitslosen von Marienthal: Ein soziographischer Versuch*. Suhrkamp, Frankfurt am Main, 9 edition, 1975.

Johnson, R. Antecendents and Outcomes of Corporate Refocusing. *Journal of Management*, 22(3):439–483, 1996.

Keim, W. Kunststoffe. Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA, 2006.

Kelly, D. und Amburgey, T. Organizational Inertia and Momentum: A Dynamic Model of Strategic Change. *The Academy of Management Journal*, 34 (3):591–612, 1991.

Kieser, A. und Woywode, M. Evolutionstheoretische Ansätze. In *Organisationstheorien*, pages 309–352. W. Kohlhammer, Stuttgart, 2006.

Kieser, A., Hegele, C., und Klimmer, M. *Kommunikation im organisatorischen Wandel*. Schäfer-Pöschel, Stuttgart, 1998.

Kirsch, A., Esser, W., und Gabele, E. *Das Management des geplanten Wandels von Organisationen*. Pöschel, Stuttgart, 1979.

Kirsch, W. *Die Handhabung von Entscheidungsproblemen. Einführung in die Theorie des Entscheidungsprozesses*. München, 3 edition, 1988.

Kirsch, W. *Unternehmenspolitik und strategische Unternehmensführung*. Kirsch, München, 1990.

Kirsch, W. *Kommunikatives Handeln, Autopoiese, Rationalität: Sondierung zu einer evolutionären Führungslehre*. Kirsch, München, 1992.

Kirsch, W. *Wegweiser zur Konstruktion einer evolutionären Theorie der strategischen Führung*. Kirsch, München, 1996.

Kirsch, W. *Strategisches Management : die geplante Evolution von Unternehmen*. Kirsch, München, 1997a.

Kirsch, W. *Kommunikatives Handeln, Autopoiese, Rationalität: Sondierungen zu einer evolutionären Führungslehre*. Kirsch, München, 2 edition, 1997b.

Kirsch, W. und Knyphausen-Aufseß, D. Z. Strategische Unternehmensführung. In *Ergebnisse empirischer betriebswirtschaftlicher Forschung : zu einer Realtheorie der Unternehmung; Festschrift für Eberhard Witte*, pages 83–114. Schäffer-Pöschel, Stuttgart, 1993.

Knyphausen-Aufseß, D. *Theorie der strategischen Unternehmensführung. State of the Art und neue Perspektiven*. Gabler Verlag, Wiesbaden, 1995.

Koch, A. *Change-Kommunikation. Erfolgskriterien für unternehmensinterne Kommunikation bei Veränderungsprozessen*. Tectum, Marburg, 2004.

Koponen, J. und Heikki, A. Strategic Management of Declining Industries - A Literature Review, 2009. URL http://eprints.herce.fi/200/. Zugriff am 22.5.2014.

Kreikebaum, H., Gilber, D., und Behnam, M. *Strategisches Management*. Kohlhammer, Stuttgart, 7 edition, 2011.

Krüger, W. Transformations-Management. Grundlagen, Strategien. In *Unternehmerischer Wandel: Konzepte zur organisatorischen Erneuerung*, pages 199–228. Gabler, Wiesbaden, 1994.

Krüger, W. Wandlungsbedarf, Wandlungsbereitschaft und Wandlungsfähigkeit als Koordinaten des Wandels. In *Excellence in Change - Wege zur strategischen Erneuerung*. Gabler, Wiesbaden, 2 edition, 2002.

Krüger, W. Management des Wandels (Change Management). In *Handwörterbuch Unternehmensführung und Organisation*, pages 1605–1614. Schäffer-Pöschel, 2004.

Krystek, U. und Moldenhauer, R. *Handbuch Krisen- und Restrukturierungsmanagement:generelle Konzepte, Spezialprobleme, Praxisberichte*. Kohlhammer, Stuttgart, 2007.

Kutschker, M. und Schmid, S. *Internationales Management*. Oldernbourg, München, 7 edition, 2011.

Lamnek, S. *Qualitative Sozialforschung*. Beltz, Weinheim, Basel, 4 edition, 2005.

Leonard-Barton, D. Core capabilities and core rigidities: A paradox in managing new product development. *Strategic Management Journal*, 13(1): 111–125, 2007.

Levitt, B. und March, J. Organizational learning. *Annual Review of Sociology*, 14:319–340, 1988.

Levy, A. und Merry, U. *Organizazional Transformation. Approaches, strategies, theories*. Praeger, New York, 1986.

Lewin, K. *Field theory in social science: selected theoretical papers*. Harper, 1951.

Locke, K. *Grounded Theory in Management Research*. SAGE, London et al., 2001.

Mahoney, J. Path dependence in historical sociology. *Theory and society*, 29 (4):507–548, 2000.

March, J. Exploration and Exploitation in Organizational Learning. *Organization Science*, 2(1):71–87, 1991.

Markides, C. Corporate Refocusing. *Business Strategy Spring*, 4(1):1–15, 1993.

Martin, J. und Eisenhardt, K. Coping with Decline in Dynamic Markets: Corporate Entrepreneurship and Recombinative Organizational Form. *Advances in Strategic Management*, 21:357–382, 2004.

Mayer, H. *Interview und schriftliche Befragung - Entwicklung, Ausführung, Auswertung*. Oldenburg, München, 4 edition, 2008.

Mayring, P. *Einführung in die qualitative Sozialforschung*. Psychologie Verlags Union, München, 1990.

Mayring, P. *Einführung in die qualitative Sozialforschung*. Beltz, 5 edition, 2002.

Mayring, P. *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. Beltz, Weinheim, Basel, 2003.

McGranth, R. G. Creating and transforming core businesses: A competence oriented analysis. Vancouver: Wharton School, Snider Entrepreneurial Center, 1995.

McKinsey. Über McKinsey. McKinsey&Company, 2014. URL http://www.mckinsey.de/ueber_uns. Zugriff am 04.09.2013.

Miller, D. Environmental fit versus internal fit. *Organization Science*, 31(2): 280–308, 1988.

Miller, D. und Friesen, P. Archetypes of organizational transformation. *Administrative Science Quaterly*, 25(2):269–299, 1980.

Mintzberg, H. Strategy formulation as a historical process. *International Studies of Management and Organization*, pages 28–40, 1977.

Mintzberg, H. Strategy formulation in an anhocracy. *Administrative Science*

Quaterly, 30(2):160–197, 1985.

Mintzberg, H. Strategy formation. School of thought. In *Perspectives on strategic Management*, pages 105–235. Harper Business, 1990.

Mintzberg, H. und Waters, J. Of Strategies, Deliberate and Emergent. *Strategic Management Journal*, 6(3):257–272, 1985.

Mintzberg, M., Ahlstrand, B., und Lampel, J. *Strategy Safary*. Free Press, New York, 1998.

Mitchell, W. The dynamics of evolving markets: The effects of business sales and age on dissolutions and divestures. *Administrative Science Quarterly*, 39(4):575–602, 1994.

Mühlbach, J. *Nachhaltigkeit von strategischem Wandel - eine systemische Perspektive*. Lang, Frankfurt, 2003.

Müller, R. *Krisenmanagement in der Unternehmung: Vorgehen, Maßnahmen und Organisation*. Lang, Frankfurt am Main et al, 2 edition, 1986.

Müller-Stewens, G. und Lecher, C. *Strategisches Management. Wie strategische Initiativen zum Wandel führen*. Schäffer-Poeschel, Stuttgart, 2001.

Nähter, C. *Erfolgsmaßstäbe der strategischen Unternehmensführung*. Kirsch, München, 1993.

Niemeier, N. *Organisatorischer Wandel aus Sicht der Synergetik*. Deutscher Universitätsverlag, Wiesbaden, 2000.

Ossadnik, W. Erscheinungsformen und Einflußfaktoren internationaler Desinvestitionen. In *Handbuch Internationales Management – Grundlagen, Instrumente, Perspektiven*, pages 991–1006. Gabler, Wiesbaden, 2002.

Perich, R. *Unternehmungsdynamik: Zur Entwicklungsfähigkeit von Organisationen aus zeitlich-dynamischer Sicht*. Haupt, Bern et al., 1992.

Peters, T. und Waterman, R. *In Search of Excellence: Lessons from Americas Best Run Companies*. Harper & Row, London, 1982.

Petraf, M. The Cornerstones of Competitive Advantage: A Resource-Based View. *Strategic Management Journal*, 14:179–191, 1993.

Pettigrew, A. Longitudinal Field Research on Change, Theory, & Practise. *Organization Science*, 1(3):267–292, 1990.

Pettigrew, A. The Character and Significance of Strategy Process Research. *Strategic Management Journal*, 13(2):5–16, 1992.

Pfeffer, J. und Salancik, G. *The external control of organizations*. Stanford University Press, New York, 1978.

PlastiCo OHG. Gesellschaftsvertrag PlastiCo OHG. Eplingen: Firmenarchiv PlastiCo, 1907.

Porter, M. *Wettbewerbsstrategie: Methoden zur Analyse von Branchen und Konkurrenten*. Campus, 13 edition, 2008.

Powell, W. W. Expanding the Scope of Institutional Analysis. In Powell, W. W. und Dimaggio, P. J., editors, *The New Institutionalism in Organizational Analysis*, pages 183–203. University of Chicago Press, 1991.

Prahalad, C. und Bettis, R. The dominant logic: A new linkage between diversity and performance. *Strategic Management Journal*, 7(6):485–501, 1986.

Prahalad, C. und Hamel, G. *Competing for the Future*. Harvard Budiness Press, 1994.

Pralahad, C. und Hamel, G. The core competence of the corporation. *Harvard Business Review*, 68(3):79–91, 1990.

Rajagopalan, N. und Spreitzer, G. Towards a Theory of Strategic Change: A Multi-Lens Perspective and Integrative Framework. *Academy of Management Review*, 22(1):48–79, 1996.

Roseboom, C. *Internationales strategisches Marketing mittelständischer Unternehmen: Eine empirische Analyse am Beispiel des Eintritts deutscher Unternehmen in den Markt Großbritannien*. Shaker, Aachen, 2004.

Schäcke, M. *Pfadabhängigkeit in Organisationen.: Ursache für Widerstände bei Reorganisationsprojekten*. Duncker & Humblot, Berlin, 2006.

Scheele, B. und Groeben, N. *Dialog-Konsens-Methodik im Forschungsprogramm Subjektive Theorien*. Francke, Tübingen, 1988.

Schendel, D. Introduction to the Summer 1992 Special Issue on "Strategy Process Research". *Strategic Management Journal*, 13(S1):1–4, 1992a.

Schendel, D. Introduction to the Winter 1992 Special Issue on "Fundamental Themes in Strategy Process Research". *Strategic Management Journal*, 13 (S2):1–33, 1992b.

Schendel, D. und Cool, K. Development of the Strategic Management Field; Some Accomplishments and Challenges. *Strategic Management Frontiers*, 17:17–33, 1986.

Schreyögg, G. *Organisation: Grundlagen moderner Organisationsgestaltung*. Gabler, Wiesbaden, 2 edition, 1998.

Schreyögg, G. *Organisation. Grundlagen moderner Organisationsgestaltung*. Gabler Springer, Wiesbaden, 4 edition, 2006.

Schreyögg, G. und Sydow, J. *Strategische Prozesse und Pfade*. Springer Gabler, Wiesbaden, 2003.

Schreyögg, G., Sydow, J., und Koch, J. *Organisatorische Pfade: Von der Pfadabhängigkeit zur Pfadreaktion?* Gabler, Wiesbaden, 2003.

Sitka, T. und Kappacher, J. Halbzeugverbunde: Hygienische Sanitäroberflächen. *Kunststoffe Fachzeitschrift*, pages 91–93, Juli 2003.

Snow, C. Reflections on European Strategic Strategic Management Research. *Strategic Management Research*, pages 307–318, 1986.

Snow, C. und Hambrick, D. Measuring organizational strategies: Some theoretical and methodological problems. *The Academy of Management Review*, 5(4):527–538, 1980.

Spiegel. Muße im Nassen. Spiegel, 1968. URL http://www.spiegel.de/spiegel/print/d-46122835.html. Zugriff am 14.08.2014.

Staehle, W. *Management: Eine verhaltenswissenschaftliche Perspektive*. Vahlen, München, 8 edition, 1999.

Stienemann, M. *Wertsteigerung durch Desinvestitioinen*. PhD thesis, Universität Göttingen, 2003.

Strauss, A. und Corbin, J. *Grounded Theory: Grundlagen Qualitativer Sozialforschung*. Beltz Psychologie Verlags Union, Weinheim, 1996.

Strauss, A. *Grundlagen qualitativer Sozialforschung - Datenanalyse und Theoriebildung in der empirischen soziologischen Forschung*. Fink, München, 1991.

Sydow, J. *Strategische Netzwerke*. Gabler, Wiesbaden, 1992.

Sydow, J., Schreyögg, G., und Koch, J. Organizational paths: Path dependency and beyond. In *Paper presented at 21st EGOS Colloquium*, number 21, jun 30 2005.

Sydow, J., Schreyögg, G., und Koch, J. Organizational Path Dependence: Opening the black box. *Academy of Management Review*, 34:689–709, 2009.

Teece, D., Pisano, G., und Shuen, A. Dynamic Capabilities and Strategic Management. *Strategic Management Journal*, 18(7):509–533, 1997.

Tripsas, M. und Gavetti, G. Capabilities, cognition, and inertia: Evidence from digital printing. *Strategic Management Journal*, 21(10/11):1147–1164, 2000.

Tushman, M. und Romanelli, E. Organizational evolution. A mehamorphosis model of convergence and reorientation. *Research in Organizational Behavior*, 7:171–222, 1985.

Tushman, M. und Romanelli, E. Organizational Transformation as Punctuated Equilibrium: An Empirical Test. *The Academy of Management Journal*, 37(5):1141–1166, 1994.

Vahs, D. *Organisation. Einführung in die Organisationstheorie- und praxis*. Schäffer-Poeschel, Stuttgart, 6 edition, 2007.

Vahs, D. und Weiand, A. *Workbook Change Mangement. Methoden und Techniken*. Schäffer-Poeschel, Stuttgart, 2010.

Venkatraman, N. The concept of fit in strategic research. Toward verbal and statistical correspondence. *Academy of Management Review*, 14(3):423–444, 1989.

Venkatraman, N. und Camillus, J. Exploring the concept of "fit" in strategic management. *Academy of Management Review*, 9(3):513–525, 1984.

Walsh, J. Managerial and organizational cognition. Notes from a trip down memory lane. *Organizational Science*, 6(3):280–321, 1995.

Waterman, R. *The renewal factor*. Bantam Books, London, 1988.

Welge, M. und Al-Laham, A. *Strategisches Management. Grundlagen, Prozess, Implementierung*. Gabler Springer, Wiesbaden, 5 edition, 2008.

Wernerfelt, B. The Resource-Based View of the Firm. *Strategic Management Journal*, 5:171–180, 1984.

Wernerfelt, B. The Resource-Based View of the Firm: Ten Years After. *Strategic Management Journal*, 16:171–174, 1995.

Wienand, U. Erfolgspotentialplanung. *Handwörterbuch der Planung*, pages 440–452, 1982.

Winter, S. Understanding dynamic capabilities. *Strategic Management Journal*, 24(10):881–995, 2003.

Witzel, A. Das problemzentrierte Interview. Forum Qualitative Sozialforschung, Online-Journal Nr. 1, 2000. URL http://www.qualitative-research.net/index.php/fqs/article/view/1132/2519.%5D. Zugriff am 21 .06.2014.

Wrona, T. *Die Fallstudienanalyse als wissenschaftliche Forschungsmethode*. ESCP-EAP, Europ. Wirtschaftshochsch., 2005.

Yin, R. *Case Study Research: Design and Methods*. SAGE, 4 edition, 2009.

Zahn, E. Strategiekompetenz – Voraussetzung für maßgeschneiderte Strategien. Stuttgart: Hrsg. E. Zahn und S. Foschiani, 1999.

Junior Management Science 1 (2016) 1-300

Junior Management Science

journal homepage: www.jums.academy

Relative Performance Information and Financial Incentives in Multidimensional Task Settings – A Conceptual and Experimental Analysis of Effects on Performance and Attention Towards Task Dimensions

Nadine Eichhorn
Ludwig-Maximilians-Universität München

Abstract

The provision of Relative Performance Information (RPI) is commonly used by firms to increase the performance of their employees. In case employees have to fulfil tasks that involve multiple dimensions, firms have to decide on the dimensionality of RPI and can basically choose between unidimensional and multidimensional Performance Information. I discuss behavioural effects of unidimensional and multidimensional RPI under different compensation schemes and apply a controlled laboratory experiment to empirically test the influence of the dimensionality of RPI on performance and attention towards task dimensions. The study demonstrates that solely the provision of unidimensional RPI improves performance in multidimensional task settings while the effects for multidimensional RPI are small and insignificant. Importantly, both unidimensional and multidimensional RPI imply a negative effect on the performance of participants with a low position in the provided ranking on overall performance. Additionally, the application of performance-based compensation negatively moderates the performance impact of both forms of RPI which though seem to be especially critical for multidimensional RPI. In this regard, multidimensional RPI induces a significant performance decrease compared to unidimensional RPI in the presence of a performance-based contract. The findings provide some indication that a distortion of attention toward the 'quantity' dimension of the applied experimental task may have caused the negative performance effect. Notably, the experimental results do not indicate increased learning effects regarding a task-specific strategy under multidimensional as compared to unidimensional RPI. Overall, no significant differences in attention towards task dimensions between the two forms of RPI can be shown.

Keywords: Relative performance information, Incentive schemes, Multidimensional tasks, Social comparison theory, Competition, Feedback

1. Introduction

Firms frequently provide relative performance information (RPI) to their employees in order to show them how well they are performing in comparison to their peers. The general intuition is that such competitive feedback can increase employee motivation and performance even when it is not linked to monetary rewards.[1] In recent years, RPI has played a fundamental role in firms' attempts to improve the productivity of their workforce through the 'gamification' of tasks.[2] This technique refers to the use of 'game design elements in non-game contexts'.[3] For instance, employees can earn points for the completion of tasks while their respective scores are compared with that of their peers in public performance rankings on virtual leaderboards.[4]

Performance comparisons between employees can be potentially based on several performance metrics since many jobs in today's business world either involve multiple tasks within a single job description or single tasks with multiple dimensions.[5] Exemplary for the latter are customer care jobs in hotel reservation centre where employees may

[1] Cf. Reeves and Read (2013); Silverman (2011); Nordstrom et al. (1991).

[2] Cf. Werbach and Hunter (2012); Reeves and Read (2013); The Economist (2012a).

[3] Deterding et al. (2011), p. 9

[4] Cf. Meder et al. (2013), p. 490.

[5] Cf. Colombino et al. (2012), p. 294; The Economist (2012b); Brown et al. (2014), p. 240.

DOI: http://dx.doi.org/10.5282/jums/v1i1pp100-137

be ranked according to the performance dimensions 'number of bookings handled' as well as 'average value of handled bookings'.[6] As indicated by this case, multidimensional task settings normally confront employees with a trade-off in attention allocation since attention spend on one task dimension cannot simultaneously expended toward another.[7] Clearly, the more time employees in a hotel reservation centre spend on each customer in order to increase the value of the booking, the lower is the number of bookings they can handle in a period of time. This implies that performance rankings on different performance dimensions in multidimensional task settings are interdependent.

Multidimensional tasks, like they are relevant in hotel reservation centre, are present across various industries and functions. In the banking sector, for example, important output dimensions of tellers' performance are 'number of new accounts opened' as well as 'average amount of funds in these accounts'. Also, in the field of software development, online programming contests that are organized by firms like TopCoder Inc. measure the number of problems accurately solved by the developer in a given time period but differentiate point values according to the chosen problem difficulty.[8] An important characteristic of the described multidimensional task settings is that performance in the different task dimensions can be aggregated into one overall performance measure. This performance measure may be for instance the total expected revenue that an employee in a hotel reservation centre generates as well as the total funds in the new accounts opened by a banks' teller or the productivity of a developer in a programming contest. In this respect, various combinations of dimension-specific performance can imply a certain level of overall performance. Therefore, as long as firms are only interested in the maximization of overall performance, they have per se no preference on how employees allocate attention towards the different task dimensions. However, since employees may differ in their abilities to perform various task dimensions, performance can significantly depend on an appropriate attention allocation. Thus, concerning the provision of RPI in multidimensional task setting, potential effects on employees' alignment between attention allocation and individual abilities have to be considered.

With respect to the characterized multidimensional task setting, firms have the opportunity to provide their employees with unidimensional or multidimensional RPI. Whereas unidimensional RPI compares peer performance on the overall performance measure that aggregates the different task dimensions, multidimensional RPI additionally includes comparative information regarding the dimension-specific performance. Based on this differentiation, it arises the question if and how the effects of RPI on performance and attention allocation are determined by its dimensionality. In this regard, prior literature on RPI only offers limited insights into its behavioural effects in multidimensional task environments. Although several experimental studies have generated informative findings on the performance impact of RPI, these mainly relate to task settings with one performance dimension and therefore do not consider a potential influence of RPI on attention allocation.[9]

Recent work of Hannan et al. (2013) indicates that prior findings on the performance effects of RPI cannot be simply generalized to a task environment with multiple performance dimensions. Specifically, the study suggests that employees which have to work on multiple tasks may start shifting their attention to one task when they are provided with task-specific RPI. Since the study is concerned with a setting in which a firm prefers equal effort on the tasks, the results imply that the provision of RPI on multiple tasks can impede overall performance improvements although employees are motivated to exert higher effort. Therewith, the findings emphasize the importance of research on the behavioural effects of RPI in task settings which imply that performance is not only determined by the motivated level of effort but also by the induced attention towards different performance dimensions. However, the work of Hannan et al. (2013) does not apply to the effects of RPI in situations where performance is not a matter of firm-preferred allocations, but is rather determined by the extent to which employees' attention towards task dimensions corresponds to their individual abilities.

Against this background, the goal of this study is to explore how the presence and the dimensionality of RPI affect performance and attention towards task dimensions in multidimensional task settings. In particular, the study aims to compare the effects of unidimensional RPI with that of multidimensional RPI. Following to insights of prior research, the effects are studied under different compensation schemes. Even when financial incentives are not tied to relative performance, they may still have a moderating influence on the behavioural effects of RPI. For instance, Tafkov (2013) indicates that the motivational effects of RPI are greater under an individual performance-based contract than under a flat-wage contract.[10] Moreover, research investigating the design of incentive contracts in multidimensional task environments suggest that financial rewards influence an employee's decision on how to allocate attention between different task dimensions.[11] Therefore, the study compares the effects of unidimensional and multidimensional RPI under flat wage compensation as well as under a performance-based contract which links financial rewards to overall task performance.

In order to address the outlined research interest, chapter 2 firstly analyses behavioural effects of RPI theoretically and presents related empirical findings. Specifically, motivational as well as learning effects of RPI may affect performance and attention allocation and are therefore examined in general

[6] Cf. Berger et al. (2011), p. 5.

[7] Cf. Hannan et al. (2013), p. 554; Holmstrom and Milgrom (1991), p. 25.

[8] Boudreau et al. (2012), p. 4-6; TopCoder (2014).

[9] E.g. Hannan et al. (2008); Azmat and Iriberri (2010); Tran and Zeckhauser (2012); Tafkov (2013).

[10] Cf. Tafkov (2013).

[11] Cf. Holmstrom and Milgrom (1991); Feltham and Xie (1994); Kachelmeier et al. (2008).

and with regard to multidimensional task settings. In addition, the influence of different compensation schemes on the motivational effects of RPI is considered. In the following, chapter 3 derives hypotheses and research questions which are tested within a controlled laboratory experiment. Chapter 4 details the methodical approach by outlining the characteristics of experimental research in general and explaining the applied experimental design as well as the experimental procedures. Thereafter, the results of the conducted laboratory experiment are reported and discussed in chapter 5. Finally, chapter 6 concludes the study and refers to its possible limitations as well as to opportunities for future research.

2. Behavioural effects of Relative Performance Information (RPI)

2.1. Motivational effects of RPI

2.1.1. Motivational effects of RPI in general

From a social-psychological perspective, the provision of RPI relates to the fundamental human social interaction process of social comparison – RPI represents social comparison information which enable a comparison between the self and others.[12] A theory of social comparison processes was first proposed by Festinger (1954) in succession of his theory of informal social communication, which emphasizes the influence of social groups on the formation of opinions. According to social comparison theory, individuals use others not only to fulfil their internal need to evaluate their opinions but also to gain knowledge about their own abilities.[13] In his seminal theory, Festinger (1954) hypothesizes that the internal need for social comparison is motivated by the basic drive to maintain an accurate self-evaluation. Since clear objective standards for such evaluations are not always available, accurate assessments of one's ability rely on the comparison with other persons.[14]

The social comparison processes induced by the provision of RPI are of particular relevance for the current analysis since they are expected to increase an individual's motivation to successfully perform a certain task.[15] Festinger (1954) states that engagement in social comparison evokes a 'drive upward' that implies an 'undirectional push to do better and better'[16] which interrelates with the aim of reducing discrepancies between oneself relative to others. These two forms of pressure manifest in competitive behaviour described as 'action to protect one's superiority'[17] which is associated with higher effort and increased performance. Therefore, related research constitute that social comparison can not only serve self-evaluation but also the motives of self-improvement as well as self-enhancement which reflect the desire to improve oneself and to protect or enhance one's self-esteem.[18]

Against this background and with respect to the large body of research that has evolved on social comparison, a more in-depth analysis is required in order to understand how the provision of RPI can affect motivation through social comparison. First, it is of interest to whom people compare themselves so as to examine requirements and characteristics of social comparison. Originally, the theory of Festinger (1954) poses that social comparison involvement requires that comparison standards are similar regarding the critical dimension. This implies that people do not compare themselves to others that have too divergent abilities.[19] Subsequent research added the factor of the similarity on related attributes such as age or educational background. Related attributes influence the performance on the critical dimension and therefore performance differences could be ascribed to them rather than to the ability on the critical dimension. Overall, the selection of similar standards is required to fulfil the motivational aspect of accurate self-evaluation.[20]

Another more detailed aspect of the question to whom people compare themselves is the direction of comparison which also addresses an individual's level of aspiration in a social comparison process. In this context, the social comparison literature fundamentally distinguishes between upward and downward comparison.[21] Based on his central tenet of a 'drive upward', Festinger (1954) predicts that people compare themselves upwards with others who are of slightly better ability.[22] Wheeler (1966) strengthens the concept of upward comparison by introducing the rank order paradigm, stating that people compare themselves in a performance ranking to those ones who rank most similar yet higher. Concerning the motives of social comparison, upward comparison is directed towards the fulfilment of the need for self-improvement as it 'can provide hope and inspiration'.[23] However, it may also negatively affect self-enhancement and the related motive of self-evaluation maintenance.[24] On the basis of this argument, research by Brickman and Bulman

[12] Cf. Guimond (2006), p. 13; Corcoran et al. (2011), p. 119.
[13] Cf. Festinger (1954); Festinger (1950); Suls and Wheeler (2000).
[14] Cf. Festinger (1954), p. 119.
[15] It should be noted that alternative theoretical approaches exist which may serve to predict a positive effect of RPI on individuals' motivation. Major examples are the following: 1) Based on goal- setting theory (cf. Locke and Latham (1990)), the provision of RPI increases motivation as it induces goal setting (e.g. achieving a high position in a performance ranking. 2) Preference-based theories hypothesize that individuals have competitive preferences and therefore derive utility or disutility from the provision of RPI as it informs them that their performance is superior or inferior to that of their peers (cf. Charness and Rabin (2002); Charness and Grosskopf (2001); Kandel and Lazear (1992)). In this context, also status-seeking preferences are of relevance which imply that status concerns for a higher rank can drive motivation under the provision of RPI (e.g. Congleton (1989); Moldovanu and amd X. Shi (2007)).
[16] Festinger (1954), p. 125.
[17] Festinger (1954), p. 126.
[18] Cf. Wood (1989), p. 232; Brown et al. (2007), pp. 60-61.
[19] Cf. Festinger (1954), p. 120.
[20] Cf. Wheeler (1966); Goethals and Darley (1977); Corcoran et al. (2011).
[21] Cf. Wheeler (1966); Garcia and Tor (2007); Brown et al. (2007).
[22] Cf. Festinger (1954), pp. 126-127; Wheeler (1966), p. 27.
[23] Cf. Suls et al. (2002), p. 161.
[24] Cf. Corcoran et al. (2011), p. 124; Tesser (1988) introduces the self-evaluation maintenance model which hypothesizes that individuals strive for the maintenance or increase of self- evaluation.

(1977) bring up the idea of downward comparison, arguing that comparisons with better performing counterparts are avoided because of the implied threatening to the self-view.[25] Wills (1981) and Taylor et al. (1983) further develop this idea and describe downward comparison as a process which enables people to increase their subjective well-being through comparisons with 'lower status targets'. In contrast to upward comparison, it reflects a process aimed to serve the motives of self-enhancement and self-evaluation maintenance.[26]

Based on the differentiation between upward and downward comparison, the question arises of how the direction of social comparison influences motivation. As past research found out, there is no clear answer to it because the affective outcome of a comparison does not depend on its direction but rather on its salient implication. More precisely, a comparison can imply an assimilative or a contrastive evaluation of the comparison target. Assimilation refers to the belief that one could achieve the same status as the comparison target whereas a contrast in social comparison emphasizes the separation between oneself and the target.[27] Therefore, an assimilation is likely to inspire an individual and to increase motivation in an upward comparison, but can on the other hand cause demotivation in case of a downward comparison. A contrast, however, has the opposite effect in upward comparisons because it may foster negative feelings of the own inferiority and lack of ability. In downward comparison, it engenders positive feelings of relief which can spur the motivation for further performance improvement. Overall, assimilation effects are expected to be predominant in comparisons because large differences between the self and the other typically lead to the cessation of social comparison.[28] Nevertheless, social comparison often entails moderate differences and prior research has identified various factors that influence whether assimilative or contrastive evaluations are implied by a comparison.[29] Commonly, it is stated that assimilation is facilitated by feelings of 'we-ness' through the identification and psychological closeness to the comparison target. Consequently, contrast is promoted by contrary feelings of 'I-ness' which are determined by factors stimulating the salience of a personal self in opposition to a social self.[30]

In his classic paper on the theory of social comparison processes, Festinger (1954) assumes that discrepancies in abilities between oneself and a comparison target result in a 'pressure toward uniformity', which implies assimilation.[31] Building on this assumption, individuals that engage in upward social comparison are expected to behave competitively based on their motivation to improve oneself as well as their desire to maintain a positive self-image and to protect self-esteem.[32] However, research has identified several variables that impact the relationship between upward social comparison and competitive behaviour. Historically, three main variables are emphasized that are closely related to the factors that facilitate assimilation. First, competitive behaviour can be intensified through the commensurability of the comparison which refers to the degree of similarity on related attributes between oneself and the comparison target.[33] Second, the psychological closeness to the comparison counterpart typically fosters competitive behaviour as it confronts an individual with a greater threat to self-evaluation maintenance.[34] Third, the higher the relevance of a performance dimension to an individual's self-definition and therefore to self-evaluation maintenance, the stronger may be the induced competitive behaviour.[35] Whereas all of these variables are highly dependent on the involved individuals, recent literature has also suggested additional situational factors that determine the relationship between social comparison and competitive behaviour. Examples include the number of counterparts in a comparison process but also the proximity of a meaningful standard for comparison, like the number one of a performance ranking or another relevant performance threshold. In this context, competitive behaviour is expected to increase as the number of comparison counterparts decreases and the proximity to a meaningful standard gets higher.[36] Overall, these findings illustrate how the strength of the motivational effects of RPI and the related competitive behaviour are determined by a variety of factors which shape the underlying social comparison.

2.1.2. Motivational effects of RPI in multidimensional task settings

So far, the analysis has applied to the basic social comparison processes and their behavioural consequences accompanying the provision of RPI in general. Building on this foundation, this section examines additional motivational effects of RPI which can arise in multidimensional task settings. As illustrated in the introduction of this study, settings in which employees have to perform tasks which include multiple performance dimensions are ubiquitous in today's business environment. Importantly, they confront employees with a trade-off in attention allocation since attention spend on one task dimension cannot simultaneously expended toward another.[37] In this regard, overall performance can significantly

[25] Cf. Brickman and Bulman (1977).

[26] Cf. Wills (1981), pp. 245-246; Taylor et al. (1983), pp. 20-21; Tesser (1988), p. 5.

[27] Cf. Suls et al. (2002), pp. 161-162; Brown et al. (2007), p. 62.

[28] Cf. Collins (2000), p. 159; Mussweiler (2003), p. 474; Festinger (1954), p. 120.

[29] E.g. Mussweiler (2001); Stapel and Koomen (2005); Collins (2000).

[30] Cf. Stapel and Koomen (2005), p. 1037; Collins (2000), pp. 159-160.; Corcoran et al. (2011), p. 128.

[31] Cf. Festinger (1954), p. 126; Corcoran et al. (2011), p. 121.

[32] Cf. Tesser (1988); Beach and Tesser (1995); Wood (1989); Brown et al. (2007).

[33] Cf. Garcia and Tor (2007), p. 98; cf. Goethals and Darley (1977).

[34] Cf. Tesser (1988), p. 5; Garcia and Tor (2007), p. 98.

[35] Cf. Festinger (1954), p. 130; Tesser (1988), p. 6; Garcia and Tor (2007), p. 98.

[36] Cf. Garcia et al. (2006); Garcia and Tor (2007); Garcia and Tor (2009); Additional situational factors ana- lysed by Poortvliet et al. (2009); Tor et al. (2013).

[37] Cf. Hannan et al. (2013), p. 554; Holmstrom and Milgrom (1991), p. 25.

depend on an appropriate attention allocation based on an individual's dimension-specific abilities. Thus, it is of special interest for this study how motivational effects induced by the use of RPI can influence an employee's decision on how to allocate attention between multiple task dimensions. Particularly, a distinction was made between unidimensional RPI that solely compares overall performance and multidimensional RPI which additionally provides comparative information on dimension-specific performance. Following this differentiation, the analysis first draws on findings on motivational effects based on self-affirmation theory and considers the social-psychological phenomenon of social distinction. It then continues the examination of the motivational impact of RPI in multidimensional task settings by analysing the influence of different competitive attitudes.

Self-affirmation tendencies and the desire for social distinction

Self-affirmation theory is built on the assumption that people are driven by a strong motivation to maintain perceived self-integrity.[38] By definition, self-integrity is 'the sense that, on the whole, one is a good and appropriate person'[39] and its evaluation is shaped by culturally or socially significant standards of integrity. Threats to an individual's perceived self-integrity can arise in different forms, including negative social comparisons which may lower self-evaluation as well as self-esteem and therefore affect self-integrity.[40] Self-affirmation theory focuses on the behavioural responses to such threats, aiming to restore a state of integrity. Overall, people can respond to a threat by the acceptance of the threatening information and corresponding behavioural changes. However, in case the threat relates to an important part of self-integrity, it might lead to defensive psychological adaptions such as the denial or the avoidance of the threat. In this context, self-affirmation theory suggests an alternative to defensive adaptations, namely an indirect psychological adaptation of affirming other resources of self-integrity.[41] This implies that people can compensate a threat to self-integrity in one area by emphasizing successes in other areas.[42]

As already stated, negative social comparison experiences are a potential source for threats toward self-integrity. Yet, based on self-affirmation theory, such a threat may be compensated by a positive social comparison in another domain to the extent that it enables the affirmation of an important self-aspect.[43] This insight has important implications for the motivational effects of RPI in multidimensional task settings. In such settings, multidimensional RPI – as opposed to unidimensional RPI – involves social comparison information which allow people to compare themselves with others on the basis of different dimensions of a task. Thereby, self-threatening social comparison in one dimension can trigger self-affirming tendencies in another dimension.[44] As a consequence, in the presence of multidimensional RPI, people may be motivated to concentrate their attention on a task dimension in which they can proof their competence. This might lead to a distortion of attention in terms of an undesirable shift of attention toward one task dimension and the neglect of other dimensions at the expense of overall performance. In contrast, unidimensional RPI limits the scope of social comparison to the overall performance measure and therefore don't provide the opportunity for self-affirmation in other dimensions.

Another social-psychological phenomenon that appears to be relevant in the analysis of motivational effects of RPI on attention towards task dimensions refers to an individual's desire for social distinction. Such an innate desire to differentiate oneselve from others is argued to be a 'hard-wired trait of human nature'.[45] Past research concerned with the motivational effects of symbolic awards highlights social distinction as a driver in award-seeking behaviour.[46] In the context of the provision of multidimensional RPI which compares people based on different task dimensions, the desire for social distinction can potentially cause motivational effects that are similar to those of a need for self-affirmation. Specifically, it may foster a selective concentration on one or more task dimensions which are seen to be possible sources of distinction.

This reasoning is supported by Snyder and Fromkin (1980) who formulate a theory of uniqueness seeking which is related to the concept of social distinction as it states that individuals strive for a moderate self-distinctiveness.[47] They constitute that similarities with others are perceived as being unpleasant and can thus trigger a motivation to positively reaffirm distinctiveness.[48] However, this assumption contradicts to some extent the prediction of social comparison theory that discrepancies in abilities result in a 'pressure toward uniformity'.[49] Thus, it is important to note that motivational effects based on the drive for social distinction as well as uniqueness seeking are – in contrast to self-affirmation theory – not induced by a negative social comparison experience. Rather, the social comparison facilitated by RPI provides a basis on which people can differentiate themselves from others and it might therefore lead toward a motivation to perform better than the comparison targets. As already described, multidimensional RPI offers various opportunities to satisfy distinction which can result in a strong focus on individual task dimensions and thereby also in a disregard of overall performance. Contrastingly, in the presence of

[38] Cf. Steele (1988); Aronson et al. (1999); Sherman and Cohen (2006).
[39] Sherman and Cohen (2006), p. 185; cf. Steele (1988).
[40] Cf. Tesser et al. (2000) for a detailed analysis of the relationship between the concepts of self-esteem and self-integrity.
[41] Cf. Sherman and Cohen (2006), pp. 186-187; Aronson et al. (1999), pp. 128-129.
[42] Cf. Sherman and Cohen (2006), p. 188.
[43] Cf. Leary and Tangney (2005), p. 287; Sherman and Cohen (2006), p. 190.
[44] Cf. Tesser et al. (2000), p. 1481.
[45] Frey (2007), p. 8.
[46] Cf. Bourdieu (1979); Ginsburgh (2003); Frey (2007).
[47] Cf. Snyder and Fromkin (1980).
[48] Cf. Lynn and Snyder (2002), p. 396; Simsek and Yalicetin (2010), p. 576; Cf. Snyder and Fromkin (1980).
[49] Festinger (1954), p. 126.

unidimensional RPI, people can only positively distinguish themselves based on their overall performance although they act in a multidimensional task setting.

Task-oriented and rivalry-oriented competitiveness

The examination of the motivational impact of RPI in general has already outlined how the presence of RPI enables social comparison which can enforce competitive behaviour described as 'action to protect one's superiority'.[50] On this basis, the consideration of socio-psychological research on competition can provide further insights regarding the motivational influence of RPI on attention towards task dimensions in multidimensional task settings.

Competition, defined in situational terms by Deutsch (1949), describes a setting in which the goal attainment of separate participants is negatively correlated, implying that a participant can only achieve his goal if the related others fail in achieving their goals.[51] With regard to this definition, RPI induces competition since it facilitates social comparison which motivates individuals to achieve a superior position in a performance ranking as compared to their counterparts. Based on the early research on competition[52], its psychological consequences and behavioural effects have been extensively studied. An important distinction underlying many of these analyses applies to the different forms of competiveness that it evokes.[53] According to Griffin-Pierson (1990), competitiveness can be defined as an achievement motive or component of achievement motivation.[54] So far, this study has not yet considered different types of competitiveness underlying the competitive behaviour which results from the provision of RPI. However, they might be a determinant of attention towards task dimensions as part of an individual's competitive behaviour. Therefore, the two main types of competitiveness which can arise from competition are discussed briefly.

The first type of competitiveness is characterized by its focus on a performance goal and the related desire to successfully perform a task and 'to be the best one can be'.[55] Various authors have described this competitiveness perspective but have labelled it differently, mainly as 'goal competitiveness'[56], 'personal development competitiveness'[57] or as 'task-orientated competitiveness'.[58] Fundamentally, this type of competitiveness refers to a traditional understanding of competition causing 'behaviour oriented toward a goal in which the other competitors for the goal are secondary'.[59] This implies that the social comparison processes induced by RPI motivate an individual primarily to increase task-related effort as well as effort directed toward learning in order to improve itself.[60] In the following, the study refers to this achievement motive as task-oriented competitiveness.

On the contrary, studies on competition also emphasize a second type of competitiveness that has its focus not on a goal but on others and is thereby called 'interpersonal competitiveness'[61], 'hypercompetitiveness'[62] or 'other-referenced competitiveness'[63]. Inherent to this type of competitiveness is an individual's desire to win and to avoid losing, accompanied with the enjoyment of interpersonal competition.[64] Thus, competition that involves this achievement motive is also framed as rivalry which is partly treated as a separate concept in past literature.[65] From the social comparison perspective, the presence of RPI can thus emphasize a motivation that is mainly about outperforming competitors as well as achieving a higher rank and not in principle on mastering a task as good as possible.[66] In this respect, an individual's needs for self-enhancement and for maintaining a positive self-image dominate the motive of self-improvement. The study labels this competiveness perspective as rivalry-oriented competitiveness.

The characterization of the two forms of competitiveness suggest that they might cause different behavioural responses to RPI in situations that offer people the opportunity to increase their position in a ranking by other means than improving overall performance. Exemplary, prior research argues that rivalry-oriented competitiveness can promote dysfunctional effects of competition like reduced reporting honesty as well as sabotage of comparison counterparts.[67] In multidimensional task settings, it might result in unintended behaviour with regard to people's attention towards the different task dimensions. In particular, multidimensional RPI enforces competition on several task dimensions which provides an individual with the chance to outrank their competitors in at least one dimension without necessarily achieving a better performance on the multidimensional task overall. Therefore, to the extent that it evokes rivalry-oriented instead of task-oriented competitiveness, multidimensional RPI might drive the motivation to focus on winning one of the competitive rankings disregarding the consequences for performance in other dimensions. This might imply a distortion of attention which can be in principle compared to the described behavioural consequences of self-affirmation and social distinction. Comparably, there is no distortion effect expected in case of the provision of unidimensional RPI because it indicates that ranking improvements can be only achieved through enhancements related

[50] Festinger (1954), p. 126.
[51] Cf. Deutsch (1949).
[52] E.g. Triplett (1898); Deutsch (1949); Johnson and Johnson (1989); May and Doob (1937).
[53] Cf. Tassi and Schneider (1997), p. 1560; Griffin-Pierson (1990), p. 108; Ryckman et al. (1996), p. 375.
[54] Cf. Griffin-Pierson (1990), p. 108.
[55] Griffin-Pierson (1990), p. 108.
[56] Cf. Griffin-Pierson (1990); Morey and Gerber (1995).
[57] Cf. Ryckman et al. (1996).
[58] Cf. Tassi and Schneider (1997).
[59] Mead (1937), p. 8.
[60] Cf. Tassi and Schneider (1997), p. 1559; Ryckman et al. (1996), p. 375.
[61] Cf. Griffin-Pierson (1990); Morey and Gerber (1995).
[62] Cf. Ryckman et al. (1990); Ryckman et al. (1996).
[63] Cf. Tassi and Schneider (1997).
[64] Cf. Ryckman et al. (1990), p. 630; Griffin-Pierson (1990), pp. 108-109.
[65] Cf. Mead (1937), p. 8; Kilduff et al. (2010), pp. 945-946; Tassi and Schneider (1997), p. 1558.
[66] Cf. Hibbard and Buhrmester (2010), p. 413; Kilduff et al. (2010), p. 961.
[67] Cf. Brown et al. (2014); Schreck (2014); Charness et al. (2013).

toward the aggregated measure of task performance.

2.2. Learning effects of RPI

2.2.1. Learning effects of RPI in general

The preceding analysis on the motivational effects of RPI is built on the finding that RPI represents an important type of information which facilitate the self-evaluation of abilities in the absence of an objective standard. Therefore, the provision of RPI serves an informational function which not only has implications for an individual's motivation but also affects its learning on how to improve performance.

According to the feedback intervention theory proposed by Kluger and Denisi (1996)[68], RPI is expected to trigger learning processes as it informs an individual about a discrepancy between its own performance and a performance standard.[69] Particularly, such a feedback-standard comparison may stimulate the search for a task-specific plan or the development of a new task-specific strategy in case an increase in effort as a 'universal strategy' proves not to be successful.[70] The learning process involves the generation and re-evaluation of working hypotheses about appropriate strategies with which one's performance can be improved. These hypotheses, in turn, also function as standards in the evaluation of the success of changed behaviour.[71] In this respect, past research on social comparison emphasizes the role of upward comparison in providing information on how to advance and make progress regarding a particular task.[72] Notably, upward comparative information can tell people 'what they would have to achieve to move up in the ability distribution'[73] and the superior counterparts might serve as role models in the learning process.[74]

However, Kluger and Denisi (1996) also point out that the effectiveness of a feedback intervention is determined by the extent to which it directs information to enhanced self-efficacy.[75] Self-efficacy can be defined as an individual's belief in its own ability to influence specific situations that affect its live.[76] The various psychological consequences of self-efficacy beliefs include effects on attentional and thinking processes. Bandura and Wood (1989) state that strong self-efficacy directs people's attention toward problem solving and the development of task strategies. On the other hand, self-doubts of efficacy may strengthen concerns on personal deficiencies as well as the concentration on possible adverse consequences of failures.[77] Consequently, feedback can not only affect an individual's learning through the information it provides about how to develop further, but also through its effect on the cognitive processes related toward the effective use of this information.

The relationship between the presence of RPI and an individual's self-efficacy beliefs depend on whether the involved social comparison implies assimilative or contrastive evaluations.[78] As described earlier, assimilation refers to an identification with the comparison target. Therefore it is suggested to raise self-efficacy in upward comparisons as it demonstrates the attainability of certain goals. On the contrary, a comparison with other from a contrastive perspective rather leads to frustration and resentment associated with lower self-efficacy.[79] Following this logic, assimilation and contrast oppositely affect perceived self-efficacy in downward comparisons. Thereby, it can be concluded that both upward assimilation as well as downward contrast in social comparison processes might increase the learning from RPI because of their positive effect on self-efficacy beliefs.

2.2.2. Learning effects of RPI in multidimensional task settings

In multidimensional task settings, learning can be related toward the improvement of dimension-specific performance as well as the refinement of the allocation of attention towards the different task dimensions. Focussing on the latter, the following analysis aims to answer the question to what extent RPI can help individuals to find an appropriate strategy on how to divide attention based on their dimension-specific abilities. For this purpose, the effects are differentiated between unidimensional and multidimensional RPI consistent with the analysis on the motivational effects of RPI.

The previous section has outlined that RPI can facilitate learning as it allows a feedback-standard comparison which is necessary for the generation and re-evaluation of working hypotheses regarding performance enhancing task-specific strategies.[80] Applied to a multidimensional task setting, the development of a strategy on the allocation of attention requires the formation of hypotheses regarding its relationship with dimension-specific performance and overall performance. In the presence of unidimensional RPI, an individual can only base these hypotheses on comparative information regarding the overall performance measure. However, this implies that an individual cannot disentangle the drivers of the performance of superior and inferior others. Especially, it does not have the opportunity to compare its strategy along with its dimension-specific abilities to that of its competitors. Therefore, an individual might generate a multitude of hypotheses since unidimensional RPI does not provide enough

[68] The feedback intervention theory by Kluger and Denisi (1996) integrates several theories and re- search paradigms on the concept of feedback. Feedback interventions are defined as 'actions taken by (an) external agent (s) to provide information regarding some aspect(s) of one's task performance' (Kluger and Denisi (1996), p. 255).

[69] Cf. Kluger and Denisi (1996), pp. 259-260.

[70] Cf. Kluger and Denisi (1996), p. 263; cf. Wood and Locke (1990).

[71] Cf. Kluger and Denisi (1996), p. 263.

[72] Cf. Corcoran et al. (2011), p. 124; Collins (1996), p. 52

[73] Collins (1996), p. 52.

[74] Cf. Lockwood et al. (2002), p. 854; Corcoran et al. (2011), p. 124; Suls et al. (2002), p. 161.

[75] Cf. Kluger and Denisi (1996), p. 276.

[76] Cf. Bandura (1977b) has established the self-efficacy theory in succession of the social learning theory (Bandura (1977a)).

[77] Cf. Bandura/Wood (1989), p. 806.

[78] Cf. Carmona et al. (2008); Suls et al. (2002).

[79] Cf. Bandura and Jourden (1991); Lockwood et al. (2002); Carmona et al. (2008).

[80] Cf. Kluger and Denisi (1996), p. 263.

information so that erroneous hypotheses can be rejected.[81] By contrast, multidimensional RPI informs an individual on how its dimension-specific and its overall performance compare with that of its counterparts. This provides the basis for the evaluation of hypotheses on the relation between dimension-specific performance, allocation of attention and overall performance. From that, inferences can be drawn with regard to the relative success of certain strategies. Consequently, multidimensional RPI appear to be more helpful than unidimensional RPI in supporting an individual's learning on how to allocate attention between task dimensions in order to improve performance.

Another concern of the preceding analysis on the learning effects of RPI in general has been the influence of RPI on self-efficacy. In general, RPI that implies upward assimilation or downward contrast can increase self-efficacy beliefs.[82] However, it arises the question if unidimensional and multidimensional RPI might differently affect self-efficacy and therewith also the effective use of the provided information for the development of task-specific strategies. Against this background, two determinants of self-efficacy can be identified that may vary depending on the informational value of RPI. Firstly, as already indicated before, self-efficacy can be enhanced by vicarious experience to the degree that it shows an individual that others can overcome difficulties in performing a task.[83] Vicarious experience rely on social comparison and has a higher impact on efficacy as it provides an individual with information on the performance strategies of superior ones.[84] In this regard, multidimensional RPI potentially can raise self-efficacy in a greater manner than unidimensional RPI. The second relevant determinant of self-efficacy is task complexity which is suggested to lower efficacy beliefs. An important component of complexity is an individual's uncertainty on its abilities for performance in situations in which the precise skills required are unclear.[85] Here, multidimensional RPI is expected to contribute to the reduction of complexity by providing social comparison information that links dimension-specific and overall performance and can therefore help to lower uncertainty on skill requirements. Overall, it appears that multidimensional RPI is more beneficial for learning processes on task strategies than unidimensional RPI not only because of the information it provides about how to allocate attention appropriately, but also through its predicted effects on self-efficacy.

2.3. Influence of financial incentives on motivational effects of RPI

2.3.1. Influence of financial incentives in general

From a standard economic perspective, the provision of RPI should not affect an individual's motivation and performance as long as compensation is not based on relative performance.[86] Therefore, the study has relied on social-psychological theories in analysing the motivational effects of RPI. This section now serves to integrate findings of economic theories in order to examine the influence of financial incentives on the predicted effects of RPI. Firstly, insights from agency theory are provided so as to analyse how financial rewards influence motivation. Afterwards, it is specifically considered how individual performance-based compensation might affect the motivational impact of RPI.

To begin with, conventional economic theories predict that a performance-based compensation contract compared to a flat wage contract motivates individuals to exert higher effort and to achieve better performance.[87] Agency theory, in particular, derives this prediction from the fundamental assumptions that individuals are fully rational and solely motivated by self-interest. On this basis, individuals are presumed to strive for a maximization of their expected utility which is described by a utility function that represents preferences for increases in wealth and leisure. Since performing a certain task involves a reduction in leisure, individuals are expected to not exert effort on that task as long as it does not contribute to their economic well-being.[88] Typically, a flat wage contract does not satisfy this condition because it implies that compensation is not linked to an individual's performance. In this context, the principal-agent branch of the agency theory is concerned with the optimal design of compensation contracts in case an individual (principal) engages another individual (agent) to perform a task on its behalf while it cannot fully observe the exerted effort. Notably, as principal-agent models commonly assume that the agents are strictly risk-averse, compensation contracts have to appropriately balance the need for motivating effort against the need for risk-sharing.[89]

Principal-agent theory refers to RPI in the context of Relative Performance Evaluation (RPE) which implies that an individual's compensation is tied to its performance relative to that of its peers. Holmstrom (1982) argues that RPE induces higher effort from risk-averse agents than individual performance-based compensation in case the agents face some common uncertainty. Since RPE allows to filter out common uncertainty from an agent's performance, it reduces an agent's exposure to 'systematic risk' without lowering its incentives.[90] Originating from this idea, past research discusses tournament compensation schemes as a specific type

[81] Cf. Kluger and Denisi (1996), pp. 263-264; Hattie and Timperley (2007), p. 93; Castellan and Swaine (1977).

[82] Cf. Bandura and Jourden (1991); Lockwood et al. (2002); Carmona et al. (2008).

[83] Cf. Bandura (1977b), p. 197.

[84] Cf. Bandura (1977b), p. 197; Gist and Mitchell (1992), p. 194.

[85] Cf. Gist and Mitchell (1992), p. 194; Cervone et al. (1986), p. 492.

[86] Cf. Holmstrom (1982); Kandel and Lazear (1992).

[87] Cf. Bonner and Sprinkle (2002) for an overview of relevant theories and of research findings on variables that moderate the relationship between financial incentives and performance (e.g. person variables, task variables, environmental variables etc.). It should be noted that conflicting social-psychological theories exist which predict negative effects of performance-based compensation, e.g. Deci and Ryan (1985) argue in their self-determination theory that financial incentives may crowd out intrinsic motivation and therefore decrease effort and performance.

[88] Cf. Baiman (1982); Baiman (1990); Eisenhardt (1989).

[89] Cf. Baiman (1990), pp. 342-343.

[90] Cf. Holmstrom (1982), p. 334-337; Antle and Smith (1986), p. 2.

of RPE. Comparably to RPI that is provided in the form of performance rankings, tournament schemes link compensation to an individual's ordinal rank in a tournament.[91]

As already mentioned, based on economic theories, RPI itself should have no effect when compensation is linked to individual performance instead of relative performance.[92] Nevertheless, considering both economic and social-psychological theories, prior research suggests that the outlined positive incentive effect of individual performance-based compensation and the predicted motivational impact of RPI have an accumulative influence on an individual's performance.[93] However, it appears to be unclear if the motivational effects of financial incentives and RPI are only additive or if there is an interaction mechanism between them. Bonner and Sprinkle (2002) conclude from a review of past literature that incentive and feedback effects in general do not interact but are typically 'independent and additive so that there is no simple two-way interaction'[94]. Yet, in a recent study, Tafkov (2013) aims to develop theory that specifically explains how individual performance-based compensation moderates the effect of RPI on performance.[95] With regard to social comparison theory, Tafkov (2013) hypothesizes that RPI is more useful for social comparison purposes under individual performance-based compensation compared to a flat wage contract. The underlying reasoning is that performance-based compensation motivates higher effort levels which results in a clearer relationship between task performance and ability. Therefore, individuals are more likely to attribute performance differences to differences in abilities which is expected to result in greater social comparison involvement and in turn to increase the motivational effect of RPI.[96] Although the study finds empirical support for this prediction, additional theoretical analyses seem to be required in order to substantiate the socio-psychological underpinnings of such an interaction mechanism. Hence, overall, there remains uncertainty whether the motivational effects of financial incentives and RPI are additive or they reinforce each other through an interaction effect.

2.3.2. Influence of financial incentives in multidimensional task settings

The analysis of the motivational impact of RPI in multidimensional task environments has shown that RPI cannot only affect effort but also peoples' attention towards task dimensions. Comparably, economic theory considers the complexities inherent in providing appropriate financial incentives in multitask and multidimensional task environments. Past literature commonly labels attention towards different tasks or task dimensions as effort allocation. In particular, agency-theoretic models are concerned with effort distortion effects which arise when compensation contracts are incomplete because multiple tasks of a job, or multiple dimensions of a task, vary in their measurability.[97] In this context, research also investigates how social pressure or social preferences like fairness influence effort distortion effects in the presence of incomplete compensation contracts.[98] However, since this study focus on settings in which all task dimensions are measurable, this stream of literature is not examined in further detail. Rather, it is of relevance how financial rewards based on an individual's overall performance influence the expected motivational effects of RPI in multidimensional task settings.

The previous section has shown that financial incentives and RPI both increase an individual's motivation to successfully perform a task which might result in an additive performance effect or even a positive interaction. In a multidimensional task setting, yet, the induced motivations might conflict with each other to the extent that they are directed towards different task dimensions. Particularly, financial incentives based on overall performance are expected to motivate an appropriate allocation of attention between different task dimensions.[99] That is, individuals should be motivated to allocate attention according to their hypothesis of how to increase overall performance based on their dimension-specific abilities. This complies with the motivational effects of unidimensional RPI but can conflict with a motivation to focus entirely on one task dimension which is potentially caused by multidimensional RPI. The latter refers to a potential distortion of attention toward a dimension in which an individual can prove its competence induced by self-affirmation tendencies, the desire for social distinction or a rivalry-oriented competitive attitude. In this context, individuals potentially face a trade-off in performance goals when provided with multidimensional RPI while compensation is tied to overall performance. It is important to note that individual performance-based compensation itself is not expected to affect self-affirmation, social distinction or rivalry as long as the monetary rewards received by competitors are not or at least not precisely known.[100] Hence, an individual's behavioural response toward such a trade-off is likely to be determined by the strength of the countervailing effects.

The motivational effects of incentive compensation are suggested to be intensified by the magnitude of a certain payoff as well as the probability of achieving it.[101] The strength of the mentioned socio-psychological processes, however, is fundamentally based on the underlying social comparison. As noted earlier, the commensurability of a comparison target, the psychological closeness toward it and the relevance

[91] Cf. Lazear and Rosen (1981); Nalebuff and Stiglitz (1983).
[92] Cf. Holmstrom (1982); Kandel and Lazear (1992).
[93] Cf. Bonner and Sprinkle (2002); Kluger and Denisi (1996); Chung and Vickery (1976).
[94] Bonner and Sprinkle (2002), p. 329.
[95] Cf. Tafkov (2013).
[96] Cf. Tafkov (2013), pp. 330-331; cf. Festinger (1954).
[97] E.g. Holmstrom and Milgrom (1991); Datar et al. (2001); Feltham and Xie (1994); Baker (2002).
[98] For example Fehr and Schmidt (2004); Brüggen and Moers (2007); Schmidt (2011).
[99] Cf. Datar et al. (2001); Baiman (1990); Prendergast (1999).
[100] Cf. Frey (2007), p. 8.; cf. Sherman and Cohen (2006); Kilduff et al. (2010).
[101] Cf. Bonner and Sprinkle (2002); Baiman (1982); Baiman (1990).

of a performance dimension reinforce the psychological effects of social comparison.[102] It is worth mentioning that rivalry-oriented competitiveness, which may be triggered by these factors, can motivate an individual to apply a task strategy that improve its position in a ranking at the cost of its economic wealth.[103] Past research associates such economically irrational behaviour with a psychological state of 'competitive arousal' which results from rivalry.[104] It implies that an individual's desire to outperform its competitors in a specific dimension of a multidimensional task can dominate its motivation to increase its wealth through a concentration on overall performance. Yet, in sum, no general conclusion can be drawn with regard to the behavioural outcomes resulting from the potentially conflicting motivations induced by financial incentives and RPI.

2.4. Empirical findings of related studies

Various experimental studies have generated empirical findings on the performance impact of RPI resulting from its behavioural effects. Based on the preceding theoretical analysis, this section reviews existing evidence on the effect of RPI on performance in general and presents empirical insights on its interaction with financial incentives. Afterwards, it outlines the findings of the few studies concerned with the effects of RPI in multitask environments since evidence from multidimensional task settings do not exist yet.

At first, studies are considered which explore performance effects of RPI based on field data. Here, recent work of Blanes i Vidal and Nossol (2011) is of relevance as it applies a data set from a firm-level quasi experiment in order to study the effect of feedback on relative performance under a piece rate pay scheme.[105] In particular, the researchers review personnel records of a German wholesale and retail organization which started to privately communicate to the workers of its main warehouse how they rank in the productivity/wage per hour distribution. The findings suggest a long-lasting 6.8 percent increase in workers' productivity following the start of the provision of the ranking information. Additionally, Blanes i Vidal and Nossol (2011) emphasize that this effect can be clearly attributed to the motivational effects of social comparison since the institutional settings implies that the ranking did not have any compensation or career consequences. In another context, research by Azmat and Iriberri (2010) also proves positive performance effects of RPI even though individuals are rewarded according to their absolute performance.[106] The study exploits data from a natural experiment that took place in a Spanish high school in which students received for only one year private information on their performance relative to the average performance in their class. The analysis shows a 5.0 percent increase in the student's grades in reaction to the provision of this information. Notably, the positive effect was significant for the whole distribution and disappeared once the information was removed.

Additional insights on the effects of RPI, however, can be drawn from the results of a field experiment analysed by Tran and Zeckhauser (2012).[107] The researchers also investigate the impact of RPI on academic performance but distinguish the effects between private and public disclosure of ranking information. Specifically, in the course of the field experiment conducted at a University in Vietnam, students enrolled in a one-semester English course were divided into three groups and either received private, public or no information on how their performance ranked relative to their peers after biweekly progress tests. Although the performance rankings brought no direct tangible benefits, the analysis reveals that the provision of RPI resulted in a higher performance of the treatment groups compared to the control group in a final test at the end of the course. Interestingly, the findings imply that the performance increases compared to the control group were around 45 percent higher for those students who received their rank publicly instead of privately.

Further evidence on the performance impact of various types of RPI are provided by laboratory experiments which also offer insights on the moderating effects of different compensation schemes. To begin with, Tafkov (2013) proves in a laboratory environment that both private and public RPI positively influence performance but that the effects are magnified when RPI is public.[108] Those findings are based on a real-effort experiment in which participants had to solve multiple-choice multiplication problems for nine independent rounds without any outside aid. At the end of rounds three, six, and nine, participants in the RPI-condition were informed about the rank of their performance within a group of five participants. In this setting, Tafkov (2013) not only manipulate the type of RPI between subjects but also test the effects of RPI under a flat-wage as well as an individual-performance-based contract. As already noted in section 2.3.1, the results support the hypothesis that the positive performance effects of RPI are stronger under performance-based compensation as it contributes to a clearer relationship between task performance and ability.

In contrast to the insights of Tafkov (2013), an experiment conducted by Murthy (2011) does not reveal significant interaction effects between RPI and different compensation schemes.[109] However, the study compares the effects of RPI under a piece-rate scheme and under two target-based schemes but does not consider a flat wage contract. The target-based schemes consisted of a budget-based scheme and quota-based scheme which both represent performance-contingent compensation contracts. Similar to Tafkov (2013), participants of the experiment of Murthy (2011) had to solve a real-effort task which, yet, involved three periods of decoding numbers into letters. After every

[102] Cf. Tesser (1988); Garcia et al. (2006); Goethals and Darley (1977).
[103] Cf. Malhotra (2010); Ku et al. (2005); Garcia et al. (2006).
[104] Cf. Malhotra (2010), p.140; Ku et al. (2005), pp. 92-93.
[105] Cf. Blanes i Vidal and Nossol (2011).
[106] Cf. Azmat and Iriberri (2010).
[107] Cf. Tran and Zeckhauser (2012).
[108] Cf. Tafkov (2013).
[109] Cf. Murthy (2011).

round, they were publicly informed about their position in a ranking of all 20 participants of the session. Overall, the results reveal a significant performance effect of RPI which is remarkably highest for participants in the bottom third of the performance distribution. Although no interaction effects are found between the provision of RPI and the different compensation schemes, the application of a quota-based contract led to significant performance enhancements which indicate an additive effect with RPI.

The review of empirical findings regarding the interaction between RPI and different compensation contracts can be expanded by considering two experimental studies which compare the performance impact of RPI under an individual incentive scheme and under a tournament scheme. First, a study of Hannan et al. (2008) examines the interaction effects between RPI and the two compensation contracts while varying the precision of RPI.[110] In particular, participants in the RPI-condition of the experiment were either provided with 'coarse' or with 'fine' feedback referring to whether RPI showed the half or the percentile of a performance distribution in which someone ranked relative to others. The task that was applied in the experiment required participants to act as managers and to make output decisions for a single product over 12 trials. The amount of profit points which they earned through their decisions were influenced by a randomly determined state of nature which induced a common error in the setting. Participants had to develop a strategy in order to effectively perform the task and they could evaluate its relative success through the RPI which was privately provided after every fourth trial. Overall, the experimental results show a disordinal interaction between RPI and the compensation scheme since RPI improved mean performance under an individual scheme regardless of its precision but negatively affected it under a tournament scheme if it was precise. The analysis indicates that this mean performance deterioration was not caused by reduced effort but rather by the adoption of ineffective strategies of those who performed below the 80th percentile in the first half of the tournament.

In addition to the insights of Hannan et al. (2008), a second study published by Eriksson et al. (2009) renders evidence on how the frequency of RPI impacts performance under the different incentive schemes.[111] In the experiment, participants were matched in pairs and had to add sets of two-digit numbers individually while they were either continuously informed about the score of their competitor or received that information halfway during the working period. Surprisingly, the results here do not comply with previous findings since they suggest that RPI does not improve performance, regardless of its frequency and the pay scheme. Instead, it is shown that the continuous provision of RPI causes subjects in both pay schemes to make more mistakes compared to when they receive no RPI. Notably, the study does not find an interaction effect between RPI and the compensation scheme. However, supplementary experimental sessions reveal that the lack of positive effects of RPI in the individual pay scheme may be due to subjects already exerting maximum effort in the absence of RPI which is why there is little room for improvements.

The presented empirical studies generate informative insights on the effects of RPI and mainly demonstrate that the provision of RPI leads to significant improvements in performance. This gives support to the expected positive influence of RPI on an individual's motivation and learning. However, the findings solely relate to unidimensional task settings with one relevant performance dimension. A first attempt to examine the generalizability of the positive performance impact of RPI to settings with multiple performance dimensions is made by Hannan et al. (2013).[112]

In particular, this study aims to capture an organizational setting in which employees have to fulfil multiple tasks with diminishing marginal returns whereby the firm has specific preferences over the allocation of effort across those tasks. Therefore, participants in the experiment had to work on both a math task and a verbal task while their compensation was not based on performance but on an equal effort allocation. RPI were provided after each of four rounds either in public or in private. Importantly, it informed participants on how they ranked in each of the two tasks among a group of five contestants. With regard to the preceding theoretical analyses, this form of RPI might induce a distortion effect in a multitask environment based on an individual's tendency to seek self-affirmation, its desire for social distinction or a rivalry-oriented competitive attitude. Accordingly, the study shows that the presence of RPI increased motivation but also induced an effort distortion effect as participants concentrated on one task while reducing their performance on the other task. Both behavioural effects were magnified under public RPI. In this context, however, the induced effort distortion was only detrimental to performance when RPI were public but did not negatively affect it under private RPI.

In a recent working paper of Hannan et al. (2014)[113], the experimental instrument used in Hannan et al. (2013) is adapted in order to examine how the effort distortion effects of public RPI depend on RPI time horizon and RPI detail. For this purpose, the study varied time horizon by providing participants with RPI that is either based on the current period performance or that refers to the cumulative performance across all periods. Additionally, RPI detail was manipulated in the form that participants either received only ordinal ranking information (less detailed) or actual performance values (more detailed) on all contestants on each task. The experimental results indicate that the effort distortion of RPI is exacerbated when RPI is more detailed or when it is cumulative. Also, an interaction effect between RPI detail and time horizon is identified since effort distorting effects of more detailed RPI are greatest when RPI is cumulative. From a theoretical perspective, the authors emphasize that these

[110] Cf. Hannan et al. (2008).
[111] Cf. Eriksson et al. (2009).
[112] Cf. Hannan et al. (2013).
[113] Cf. Hannan et al. (2014)

observed effects might be driven by underlying social comparison processes and the desire for social distinction. While cumulative RPI can be more useful for social comparison purposes as it offers a more stable and meaningful measure of ability, more detailed RPI should enable participants to differentiate themselves to a greater degree from their peers. Therefore, participants might have a greater motivation to distort their effort away from an equal allocation and to concentrate on one task in which they can proof their competence.

3. Hypotheses and research questions on effects of RPI on performance and attention towards task dimensions

The objective of this study is to explore how the presence and the dimensionality of RPI affect performance and attention towards task dimensions in multidimensional task settings. In particular, the study aims to compare the effects of unidimensional RPI with that of multidimensional RPI under different compensation schemes. To address this research interest, two directed hypotheses and four two-tailed research questions are developed building on the preceding analysis of the behavioural effects of RPI.

First, the performance effects of unidimensional RPI are considered in order to establish the basis for a comparison with multidimensional RPI. From a social-psychological perspective, unidimensional RPI facilitates social comparison by which an individual can assess its overall ability to perform a multidimensional task in the absence of an objective standard. In this context, social comparison is expected to drive an individual's motivation to enhance its performance on the overall task and to reduce performance discrepancies to superior others. These motivational effects should manifest in competitive behaviour involving higher effort and also the alignment of attention towards task dimensions to the overall performance measure.[114] Therefore, unidimensional RPI is expected to positively affect performance in multidimensional task settings. Consistent with this prediction, experimental studies reviewed in chapter 2.4 provide empirical support for performance improvements induced by RPI.[115] Even though these studies relate to unidimensional task settings, they strengthen the outlined expectation since there is no theoretical indication that motivational effects of unidimensional RPI differ in multidimensional task setting. In addition, unidimensional RPI can increase learning on task-specific strategies because of its informational value as well as its potential positive effect on self-efficacy.[116] However, section 2.2.2 emphasizes that learning effects might be limited because of the multidimensionality of the task setting. Overall, the following hypothesis formally summarizes the outlined expectations:

> *H1: Overall performance is greater when individuals receive unidimensional RPI compared to when they receive no RPI.*

Next, the performance implications of providing unidimensional RPI under different compensation schemes are considered. As illustrated in chapter 2.3.1, economic theories, such as agency theory in particular, predict that performance-based compensation motivates higher performance than a flat wage which is not tied to performance.[117] In this regard, the study aims to examine the performance effects of RPI in multidimensional task settings under an individual performance-based scheme rewarding overall performance and under a flat wage contract. According to prior research, the motivational effects of financial incentives and that of RPI generally have an accumulative impact on performance.[118] Thus, given that unidimensional RPI facilitates social comparison regarding the performance measure that is relevant for compensation, performance is expected to increase when unidimensional RPI is provided together with performance-based instead of flat wage compensation. However, as outlined in section 2.3.1, it appears to be unclear whether this will be driven by an addition or a positive interaction between the motivational effects of RPI and individual performance-based compensation. Therefore, the expectations are restricted to the performance difference between the provision of unidimensional RPI under an individual performance-based scheme and a flat wage which is formally captured in the following hypothesis:

> *H2: Overall performance is greater when individuals receive unidimensional RPI and are compensated based on an individual performance-based contract compared to when they are compensated based on a flat wage contract.*

In the following, it is considered how multidimensional RPI compared to unidimensional RPI might differently affect overall performance because of its potentially deviating influence on attention towards task dimensions. Here, it seems to be preferable to develop two-tailed research questions instead of directed hypotheses. Firstly, the theoretical analysis of the behavioural effects of multidimensional RPI in chapter 2 implies conflicting influences on attention allocation which have not been studied in this context before. Secondly, the resulting overall performance cannot be clearly predicted as it is determined by the interplay between attention allocation and individual abilities.

Multidimensional RPI might have a different influence on attention allocation since it does not limit the scope of social comparison to the overall performance measure but

[114] Cf. Festinger (1954); Suls et al. (2002); Buunk and Mussweiler (2001).
[115] E.g. Blanes i Vidal and Nossol (2011); Azmat and Iriberri (2010); Tran and Zeckhauser (2012); Tafkov (2013).
[116] Cf. Kluger and Denisi (1996); Bandura (1977b); Bandura and Jourden (1991); Corcoran et al. (2011).
[117] Cf. Baiman (1982); Baiman (1990); Eisenhardt (1989).
[118] Cf. Tafkov (2013); Bonner and Sprinkle (2002); Kluger and Denisi (1996); Chung and Vickery (1976).

also facilitates peer comparisons on dimension-specific performance. As shown in section 2.1.2, this can induce a selective concentration on a task dimension in which an individual has the chance to outperform its peers and can therefore achieve self-affirmation, social distinction or satisfy its desire to win a contest.[119]

Two experimental studies, Hannan et al. (2013) and Hannan et al. (2014), find empirical support for the described behavioural effect of RPI in a multitask environment in which individuals are provided with task-specific rankings but do not receive RPI on overall performance.[120] Importantly, multidimensional RPI also compares overall performance and therefore it may only cause differences in attention allocation to the degree that it promotes a strong concentration on one of the individual task dimensions that does not reflect overall performance. However, in contrast to this motivational effect which potentially induces a distortion of attention, multidimensional RPI also implies a learning effect that may promote an appropriate allocation of attention according to individual abilities. Particularly, section 2.2.2 has outlined that multidimensional RPI can be more beneficial for learning processes on task strategies than unidimensional RPI through its informational value and its expected greater effect on self-efficacy.[121] Therefore, both an individual's motivation and learning under multidimensional RPI can cause differences in attention allocation compared to unidimensional RPI but from a theoretical perspective. However, from a theoretical perspective, it appears to be unclear how these effects ultimately manifest in behaviour. Also, so far no empirical studies provide insights on that question since Hannan et al. (2013) and Hannan et al. (2014) exclude the influence of overall rank information from their experimental setting and do not consider learning effects. Consequently, the following research question is stated:

> *RQ1a: Will attention towards task dimensions differ when individuals receive multidimensional RPI compared to when they receive unidimensional RPI?*

Elaborating on this research question, the influence of financial incentives on attention allocation differences between unidimensional and multidimensional RPI is taken into consideration. Section 2.3.2 has illustrated that a compensation contract based on overall performance implies motivational effects which might conflict with the described social-psychological incentives for a distortion of attention under multidimensional RPI. Therefore, people can face a trade-off in attention allocation to the degree that the motivations induced by the compensation contract and by multidimensional RPI are directed towards different task dimensions. The behavioural outcome of this trade-off is expected to be determined by the strength of these potentially countervailing effects which can be influenced by various individual and situational factors.[122] In this context, it can be noticed that the introduction of financial incentives can weaken attention distortion effects caused by multidimensional RPI. However, by directing attention towards overall performance, this may in turn increase the impact of learning benefits associated with the provision of multidimensional RPI. Thus, overall it cannot be predicted to which extent attention allocation differences between unidimensional and multidimensional RPI may be present when compensation is tied to individual overall performance.

This is captured in the research question below:

> *RQ1b: When individuals are compensated based on an individual performance-based contract, will attention towards task dimensions differ when they receive multidimensional RPI compared to when they receive unidimensional RPI?*

Based on the consideration of differences in attention towards task dimensions under multidimensional RPI compared to unidimensional RPI, implications for overall performance are taken into account. Importantly, there is no indication that the two forms of RPI vary in their effect on the overall level of effort expended on a task which is why the development of research questions on potential performance differences relates to differences in attention allocation. Since individuals are expected to differ in their abilities to perform various dimensions of a multidimensional task, overall performance can significantly depend on an appropriate allocation of attention. As it was emphasized before, multidimensional RPI may foster a strong concentration on a task dimension that is different from overall performance but is perceived to offer the highest chance for an individual to prove its competence and to outperform its peers. Thus, an individual is motivated to focus attention on its relative strengths within a multidimensional task which can involve positive and negative effects on overall performance. More concretely, the performance effects are determined by the marginal return of effort implied by the allocation of attention. A shift of attention toward one specific task dimension positively affects performance as long as it increases the marginal return of overall effort spend on a task compared to an alternative allocation of attention. Yet, the extent to which the marginal return can be increased by focussing on one dimension is naturally limited in multidimensional task settings. Depending on an individual's abilities, overall performance will be negatively affected when the attention devoted toward a specific dimension reaches a certain proportion at which marginal return decreases. However, it cannot be predicted to which extent individuals shift their attention based on the motivational effects of RPI. Besides, learning effects may be also a source of differences in performance resulting from the provision of the two forms of RPI. It

[119] Cf. Steele (1988); Sherman and Cohen (2006); Frey (2007); Kilduff et al. (2010).

[120] Cf. Hannan et al. (2013); Hannan et al. (2014).

[121] Cf. Kluger and Denisi (1996); Hattie and Timperley (2007); Bandura (1977b); Cervone et al. (1986).

[122] Cf. Bonner and Sprinkle (2002); Baiman (1990); Tesser (1988); Garcia et al. (2006).

was constituted before that multidimensional RPI promotes the development of a task-strategy that aligns attention towards task dimensions with an individual's abilities. Therefore, individuals are expected to be better able to maximize their marginal return to effort under multidimensional compared to unidimensional RPI. This in turn can imply possible performance improvements. In sum, the two forms of RPI should lead to different performance to the extent that their motivational and learning effects manifest in differences in the fit between attention towards task dimensions and individual abilities. On this basis, two research questions are stated which reflect the prior differentiation between possible attention allocation differences in general and in the presence of an individual performance-based compensation contract:

> *RQ2a: Will overall performance differ when individuals receive multidimensional RPI compared to when they receive unidimensional RPI?*
>
> *RQ2b: When individuals are compensated based on an individual performance-based contract, will overall performance differ when they receive multidimensional RPI compared to when they receive unidimensional RPI?*

4. Methodological approach

4.1. Laboratory experiments as a research method to study behavioural effects of RPI

In the present study, the derived hypotheses and research questions are tested within a controlled laboratory experiment. Therefore, this section briefly outlines methodical characteristics of experimental research with regard to the examination of the behavioural effects of RPI in laboratory settings.

Experimental studies in general are characterized by an active and purposeful manipulation of an independent variable in order to measure its influence on a dependent variable. By using the principle of randomization, experiments allow to control for the influence of confounding variables and hence provide the opportunity to draw causal inferences regarding the relationship between the observed variables.[123] However, compared to field experiments that take place in natural environments, laboratory experiments create artificial situations in which researchers can investigate cause-effect relations 'under pure and uncontaminated conditions'[124]. In the laboratory, the action and decision environment of experimental participants is subject to tight control and variables can be precisely and objectively measured. Accordingly, the generated results are expected to be replicable and have a higher degree of internal validity. Yet, the abstraction and simplification involved in laboratory experiments is also accompanied by a reduction of external validity in comparison to field experiments. Therefore, the possibilities to generalise results from the laboratory context to real-word settings are limited.[125]

Based on the advantage of laboratory experiments in ensuring internal validity and establishing causal inferences, they are especially useful for testing, comparing and refining theoretical predictions and assumptions.[126] Therefore, various empirical studies employ laboratory experiments in order to examine the behavioural effects of RPI.[127] The flexibility inherent in the laboratory approach allows the researchers to vary the provision of RPI in different settings, to test boundary conditions and to isolate the influence of economic and social-psychological factors.[128] In this way, it is possible to identify precise interrelations of involved variables and to gradually expand the insights on the behavioural regularities implied by RPI. Accordingly, based on previous findings, the current study conducts a laboratory experiment in order to distinctively test its hypotheses and research questions in a controlled setting and thus to provide a first understanding of the behavioural validity of theoretical predictions concerning the presence of RPI in multidimensional task environments.

As indicated before, experimental results generated in the laboratory have various desirable characteristics but researchers also face concerns regarding their external validity. Commonly, it is argued that the findings are exposed to a subject pool bias since the experimental subjects are often students. Such objections are particularly relevant for experiments that examine individual behaviour in tasks which require professional experience like for example asset trading or product pricing. Also, student samples seem to be problematic if experiments aim to describe population parameters like risk preferences or if they are interested in the influence of socio-economic factors on certain preferences or behaviour.[129] In contrast, however, the presented experimental studies on RPI as well as the current study test theoretical predictions that relate to fundamental behavioural patterns which are independent of assumptions regarding the participant pool. Therefore, it is important to note that student samples indeed lack representativeness but that this is generally not expected to be a major threat to the validity of the experimental findings on the behavioural effects of RPI.[130]

Another concern on the external validity of laboratory experiments relates to the question if they capture essential conditions that prevail in reality. Experiments cannot integrate all conditions of a real-word environment, since abstraction and simplification are necessary in order to ensure internal validity. Essentially, it depends on the purpose of the experiment to which extent the manipulations should be applied in a realistic context. Laboratory studies on RPI typically add realism by conducting 'real effort' experiments where subjects do not choose hypothetical effort

[123] Cf. Sprinkle (2003), p. 289; Schnell et al. (2008); p. 212.
[124] Sprinkle (2003), p. 289; cf. Kerlinger and Lee (2000).
[125] Cf. Loewenstein (1999), p. 26; Schnell et al. (2008), p. 219.
[126] Cf. Smith (1994); Simon (1987).
[127] Cf. Chapter 2.4.
[128] Cf. Sprinkle (2003), p. 289.
[129] Cf. Feltovich (2011), p. 372; Falk and Fehr (2003), p. 401.
[130] Cf. Falk et al. (2009), p. 55; Falk and Heckman (2009), p. 537.

levels based on a monetary function but actually work on a specific task.[131] In this context, however, researchers have to accept some loss of control since they cannot observe the participants' cost of effort. In addition, some studies examining the impact of public RPI in comparison to private RPI required experimental subjects to introduce themselves to the others at the beginning of the session.[132] Therewith the researchers removed the anonymity in the laboratory and modelled the setting closer to a workplace context in which colleagues know each other. Nevertheless, experiments cannot implement mundane realism and should only include additional context variables when they are expected to directly enhance the validity of the results since it may also reduce the superior control possibilities in the laboratory.[133]

In sum, it is important to note that results drawn from laboratory experiments can be usefully complemented with findings from field experiments or quasi-experiments.

Building on clean evidence generated in the laboratory, field data can provide representative insights since they capture individual's behaviour in real situations and typically include larger samples.[134] Therefore, research progress regarding the behavioural effects of RPI is driven by both laboratory and field studies.

4.2. Specification of the experimental design

4.2.1. Real effort task

In accordance with previous laboratory experiments examining the behavioural effects of RPI, this study employs a 'real effort' experiment in order to test its hypotheses and research questions. The current chapter hence starts with a description of the applied 'real effort'-task which is then used a basis to explain the experimental manipulations.

Participants of the experiment are required to perform a multidimensional task on an individual computer terminal for six consecutive rounds of 360 seconds each.[135] The task consists of solving multiplication problems with different difficulty levels. In particular, participants are provided with task blocs that include six multiplication problems which are ordered according to increasing difficulty. All problems of a bloc are shown on one screen but participants can only solve them in the given order. Therefore, participants have to start each task bloc with the problem of the lowest difficulty level and can only proceed to the next level when they solve it correctly. After entering a solution to a problem, participants have to click on the corresponding 'OK'-button so that the program can check whether the answer is correct or not. A warning message appears in case participants enter wrong answers and then they have the opportunity to retry solving the problem. Importantly, participants do not have to deal with all difficulty levels included in a bloc but can also change to the next task bloc after solving the first problem of the lowest difficulty level. However, participants receive points for solving the multiplication problems and the higher the difficulty level, the more points can be earned on a problem. Therefore, the points per task bloc rise progressively with the difficulty level. Exemplary, a task bloc with two correctly solved problems accounts for 3 points while a bloc with four problems accounts for 16 points. Once participants click on the corresponding button to change to the next bloc, they are prohibited from returning to the previous one. Per round, a maximum of 75 task blocs can be displayed which is yet far beyond the number of blocs that participants can solve in 360 seconds even if they only work on problems with the lowest difficulty level. The task blocs are presented in the same order to all participants and have to be solved without any outside aid which also includes writing materials.

The total number of points earned by participants reflects their overall task performance and is determined by their performance in two task dimensions – the number of task blocs solved as well as the average number of points earned per bloc. Participants can individually decide at which difficulty level they change from one task bloc to another and therefore they have the opportunity to adopt various task strategies. Notably, there is no dominant strategy since participants' overall performance depend on the extent to which they align their attention towards the two task dimensions with their abilities in solving multiplication problems. For example, a certain participant may make the best out of his abilities by solving many task blocs with a rather low number of points per bloc. In contrast to such a 'quantity'-strategy, another participant can be able to maximize his overall performance by following a 'value'-strategy which implies that he solves fewer task blocs but concentrates on achieving a higher number of points per bloc.

Following a math task applied by Hannan et al. (2013), the differentiation between the difficulty levels of the multiplication problems is based on two criteria: (1) number of digits in the two multiplicands (e.g. multiplying a one-digit and a two-digit number compared to multiplying two, two-digit numbers) and (2) frequency of carry-overs required to solve the problem.[136] Various pre-tests have been conducted to specify the two criteria in a manner that the time required for solving a problem increases considerably with the difficulty level. Additionally, the increase in points that can be earned per difficulty level is balanced so that the marginal return of solving an additional problem differs considerably with the participants' abilities. Specifically, the time that it takes a participant to solve a problem determines the difficulty level until which there is an increase in points earned per second of time that is spend on a task bloc. Overall, this ensures that the task design is appropriate for testing the underlying theory.

[131] Cf. Falk and Fehr (2003), p. 404; v. Dijk et al. (2001).
[132] Cf. Tafkov (2013); Hannan et al. (2013).
[133] Cf. Libby et al. (2002), p. 797.
[134] Cf. Falk et al. (2009), pp. 5-6; Falk and Ichino (2006); Azmat and Iriberri (2010).
[135] The experiment was programmed using z-Tree software, cf. Fischbacher (2007).
[136] Cf. Hannan et al. (2013), p. 559-560. See Online-Appendix 1 for an overview on the specification of the difficulty levels, including the number of points that participants could earn per difficulty level.

4.2.2. Experimental manipulations

The conducted experiment is characterized by a 3 x 2 between-subjects design that implies the manipulation of two factors which represent the primary independent variables of interest. The first factor constitutes the provision of RPI and is varied a three levels: No RPI, Unidimensional RPI and Multidimensional RPI. The second factor is the compensation contract which is manipulated at two levels: Flat wage contract and individual performance-based contract. Therefore, the experiment involves six treatments which are applied between-subjects so that each of the participants is exposed to one of the treatments.

The use of a between-subject design has the advantage over a within-subjects design that carryover-effects between treatments are avoided which can otherwise threaten the internal validity of the results. Additionally, a between-subject design reduces 'demand effect' concerns which relate to the risk that participants form assumptions over an experiment's purpose and change their behaviour accordingly. Therefore, previous experiments examining the behavioural effects of RPI systematically rely on between-subjects design. However, it should be also noted that within-subjects designs in comparison have the upside that fewer participants are required to generate statistical powerful results since it is controlled for subject-specific noise.[137]

In all treatments, participants receive individual performance information at the end of each round showing them their overall score as well as the number of task blocs solved and the average number of points earned per task bloc. The latter only takes into account task blocs that are completed, meaning that participants have clicked on the corresponding button to change to the next bloc. Therewith, it is ensured that the performance measure is not distorted by unfinished task blocs at the end of the rounds. The individual performance information is provided with regard to the current period performance and the cumulative performance across all periods. During all rounds, participants can keep track of their current period performance which is displayed on the top of the task screen.

In the 'No RPI' treatment, no additional performance information is presented to the participants. In the treatments that involve the provision of RPI, participants are assigned to groups of five and receive public ranking information that compare their cumulative performance to that of the other group members at the end of each round. For that purpose, participants have to choose one out of eight given fictional nicknames at the beginning of the experiment which they keep over all rounds.[138]

Since anonymity has to be preserved in order to comply with the guidelines of the laboratory in which the experiment is conducted, participants do not know how the chosen names match to the actual persons in the laboratory. In general, the presented RPI only informs participants about performance ranks but they do not learn about the specific performance levels of the other group members. In this way, RPI cannot be used for wage comparison purposes in the treatment that applies an individual performance-based contract.

Participants in the 'Unidimensional RPI' treatment receive only one ranking comparing the cumulated number of points of the group members after every round. In the 'Multidimensional RPI' treatment, two additional rankings are provided that involve performance comparisons related to the number of task blocs solved and the average number of points earned per task bloc. In both treatments, participants with the highest rank after a certain round see a picture of a golden trophy on their computer screen next to the corresponding ranking. At the end of the experiment, congratulations to the final winners of the rankings are displayed publicly in addition to the privately shown trophies. Through the combination of RPI with such symbolic rewards, the setting is modelled closer to a 'gamification' context in which employees receive virtual batches that demonstrate their relative success in a certain task or task dimension.[139]

As indicated before, the manipulations related to the provision of RPI are crossed by the variation of the compensation contract at two levels. In the 'Flat wage' treatment, participants earn a fixed amount of 1 Euro for each of the six rounds independent of their performance on the task. In contrast, the compensation in the 'Individual performance-based' treatment is determined by the total number of points that participants have achieved at the end of the experiment. Specifically, they receive 0.2 Euro for each point which implies that they can earn a maximum of 1.10 Euro for a completely solved task bloc which accounts for 55 points.

4.3. Experimental procedures and participants

The experiment was conducted in the experimental laboratory MELESSA[140] at the University of Munich. It involved six separate experimental sessions with each representing one of the described treatments.

The sessions were divided into two independent parts – the first part was designed to elicit the participants' risk preferences and the second part represented the actual real effort experiment. Upon arrival, participants were told that they would go through two independent experiments and then received instructions on the first part which were read aloud by the experimenter.[141] Participants were not informed what the second part would involve. The measurement of the participants' risk preferences in advance to the real effort experiment is necessary in order to control for the influence of risk preferences on attention towards task dimensions. In

[137] Cf. Libby et al. (2002), p. 804; Schnell et al. (2008), p. 210; Field and Hole (2002), p. 75.

[138] The given nicknames include gender-neutral common surnames which did not allow for conclusions on personal characteristics. The members of a group are required to choose different names.

[139] Cf. Deterding et al. (2011); Reeves and Read (2013).

[140] MELESSA abbreviates 'Munich Experimental Laboratory for Economic and Social Sciences'.

[141] See Online-Appendix 2 for the instructions on the first part of the experiment.

particular, the risk involved in solving multiplication problems increases with the difficulty level of the problems since participants get more uncertain about whether they will be able to solve the problem correctly and about how long it will take. Therefore, risk aversion might be a determinant of attention allocation when compensation is tied to individual performance.[142] The experimental design that is applied to elicit participants' risk preferences is similar to one developed by Holt and Laury (2002) and consists of simple lottery choices.[143] Participants were required to make ten choices between paired lotteries and were informed that one choice will become relevant for payment by which the incentive compatibility of each decision is ensured. The information on the relevant choice as well as the associated payment were revealed at the end of the experimental session.

Upon completion of the first part, the instructions for the actual real effort experiment were handed out and read aloud.[144] Participants had the chance to ask questions and were then required to take a pre-experiment quiz. So as to ensure that participants understand the experiment, they could not proceed until they scored 100 percent on the quiz.[145] After all participants passed the quiz, they were given the opportunity to practice the task in a training round in which no manipulations were applied. The instructions emphasized that participants should use the training round in order to develop a task strategy that fits their abilities. Following to the training round, participants could again ask questions in case anything remained unclear to them. Next, participants in the treatment conditions that involved the provision of RPI had to choose a fictional name which they kept for the rest of the experiment. Then, participants performed the real effort task for six consecutive rounds. Afterwards, they were asked to complete a post-experiment questionnaire that included questions concerning controls and demographic data.[146] Finally, participants privately received their payment from the experiment in cash. On average, participants earned 12.7 Euro for approximately one hour of participation, including the payment from the lottery of averagely 2.5 Euro as well as the show-up fee of 4 Euro. Notably, the average earnings did not differ significantly between the two compensation treatments (t=0.54, p=0.58, two-tailed).

The experiment was conducted with 150 participants so that each session employed 25 participants. All participants were students that had been recruited from the participant pool of the experimental laboratory MELESSA through an anonymous invitation procedure using ORSEE software[147]. Of the participants, 86 were male (57.3 percent) and 64 were female (42.7 percent). Overall, participants were similar on related attributes like age, education and mathematical background. The mean age was 22.9 (s.d.=3.66) and the participants were primarily undergraduate students (74.7 percent)[148]. Furthermore, most of them majored in a math-related field of study (80.6% percent)[149]. With regard to their risk preferences, the results of the first experiment indicate that the participants were overall risk averse. On average, their ten lottery choices involved 5.99 (s.d.=1.64) save choices.[150]

Additionally, the participants' preference for competition is measured in the post-experimental questionnaire by a question adapted from Griffin-Pierson (1990).[151] Using a seven-point Likert scale, participants had to assess the degree to which they perform better when they are competing against someone rather than when they are the only one striving for a goal (1=not at all, 7=to a great degree). The participants' average assessment equaled 4.09 (s.d.=1.99). Furthermore, the questionnaire included control questions which required participants to indicate on the same scale the degree to which they were interested in the task (mean=4.23, s.d.=2.06) and to which they knew the task from prior experiments (mean=1.29, s.d.=0.95).

Across the treatments, there was no significant variation in the presented personal variables except for some differences in gender, age, mathematical background and risk aversion.[152] Therefore, the multivariate analysis of the experimental results controls for the influence of these individual characteristics on performance as well as on attention towards task dimensions.

5. Results

5.1. Tests of hypotheses and research questions

5.1.1. Bivariate analyses

The two primary dependent variables of the analysis of the experimental results are performance and attention towards task dimensions. The performance of the participants is measured by the total number of points that they earned through solving multiplication problems over six rounds. The

[142] Agency theory emphasizes that risk-averse individuals dislike the income risk that is associated with performance-based compensation (cf. Baiman (1990), Prendergast (1999)). In the applied set- ting, participants can reduce their income risk by concentrating on multiplication problems with low difficulty level.

[143] Cf. Holt and Laury (2002).

[144] See Online-Appendix 3 for the instructions on the real effort experiment.

[145] See Online-Appendix 4 for the questions of the pre-experiment quiz.

[146] See Online-Appendix 5 for the questions of the post-experimental questionnaire.

[147] Cf. Greiner (2004).

[148] The rest of the participants either had already completed a bachelor's degree (16.0 percent) or a master's degree/diploma (9.3 percent).

[149] The math-related fields of study include Business & Economics (55.3 percent), Natural Science & Engineering (17.3 percent) as well as Formal Sciences (8.0 percent).

[150] Cf. Holt and Laury (2002), p. 1649, 5 save choices indicate 'slight risk aversion', 6 save choices indi- cate 'risk aversion'.

[151] Cf. Griffin-Pierson (1990).

[152] 1) Gender: 40% male participants under Multidimensional RPI/Flat Wage, 64% under Unidimen- sional/Flat Wage, No RPI/Flat Wage, No RPI/Perf.-based (t=1.71, p=0.09, two-tailed). 2) Age: Mean age of 21.76 under No RPI/Perf.-based, 23.88 under Multidimensional RPI/Perf.-based (t=1.98, p=0.06, two-tailed). 3) Math Background: 68% of participants major in math-related field of study under Unidimen- sional RPI/Flat Wage, 92% under No RPI/Perf.-based (t=2.18, p=0.04, two-tailed). 4) Risk Preferences: Mean number of save choices is 5.68 under No RPI/Flat Wage, 6.64 under Unidimensional RPI/Flat Wage (t=2.10, p=0.04, two-tailed).

measure of attention towards task dimensions is defined by the participants' average difficulty level per task bloc which underlies their overall performance.[153] This variable reflects the participants' allocation of attention between the 'value' dimension and the 'quantity' dimension of the task since it indicates at which difficulty level participants have changed from one task bloc to another.[154]

Table 1 and Table 2 report summary statistics for performance as well as attention towards task dimensions and provide bivariate comparisons between treatments. In Panel A of each of the Tables, treatment data are pooled for the three levels of RPI. Panel B presents a comprehensive overview on all six treatments including the effects of the variation of the compensation contract. Figure 1 and Figure 2 each show a graphical summary of the mean values presented in Panel B of Table 1 and Table 2. Since both dependent variables do not follow a normal distribution[155], non-parametric tests are used to statistically compare treatments. First, p-values are reported for a two-tailed Mann-Whitney pairwise test which examines whether two independent samples are drawn from populations with the same distribution.[156] In order to allow for a more comprehensive analysis, the results of a Pearson Chi-Squared test on the equality of medians[157] are provided in addition because the data samples on the dependent variables both include various (extreme) outlying observations[158]. Overall, each participant is treated as an independent observation which also applies to further statistical tests that are used in the course of the analysis.[159]

In hypothesis H1, it is predicted that participants perform better under the provision of unidimensional RPI compared to when they do not receive RPI. The data in Panel A of Table 1 support this prediction since they show that performance is significantly higher under unidimensional RPI (p=0.06 / 0.05)[160]. Notably, the standard deviation of the performance of participants provided with unidimensional RPI is also considerably higher which reflects a greater spread of the data.[161] Panel B of Table 1 shows that this spread is especially driven by the treatment involving performance-based compensation since the related standard deviation of performance clearly exceeds the one that can be observed under a flat wage contract. Furthermore, in accordance with the prediction of hypothesis H2, mean values indicate a performance increase when participants receive performance-based instead of flat wage compensation in addition to unidimensional RPI. As illustrated in Figure 1, this performance effect seems to be additive since the induced performance difference between the two compensation contracts is similar when participants are not presented with RPI. However, in contrast to the performance effect suggested by the mean values, the related medians imply a slightly negative effect of performance-based compensation. This difference between mean and median values can be explained by the strongly right-skewed distribution of performance under a performance-based contract[162] which includes two extreme outliers[163]. Consequently, the p-values obtained from the comparison between performance-based and flat wage compensation under unidimensional RPI reveal that there is no significant performance difference (p=0.82 / 0.40). Therefore, in sum, the results do not support the prediction of hypothesis H2.

The research questions RQ2a and RQ2b address the performance effect of multidimensional RPI compared to that of unidimensional RPI. First, with regard to RQ2a, Panel A of Table 1 suggests that participants perform worse when they receive multidimensional instead of unidimensional RPI. Nevertheless, the reported p-values narrowly fail to confirm a significant performance difference between the two forms of RPI (p=0.19 / 0.11). Yet, participants provided with multidimensional RPI also do not perform significantly better than those in the 'No RPI' treatments although this is suggested by mean and median values (p=0.79 / 0.69)[164]. This contrasts the significant performance effect of unidimensional RPI which was presented before. Further insights on the influence of multidimensional RPI on performance can be drawn from Panel B of Table 1. Importantly, summary statistics indicate that the performance decrease under multidimensional compared to unidimensional RPI is more pronounced when participants are compensated with a performance-based contract. This interaction between the dimensionality of RPI and the applied financial incentives is graphed in Figure 1. In his context and with regard to RQ2b, the Pearson Chi-Squared test

[153] In the following, 'average difficulty level', 'difficulty level' and 'attention towards task dimensions' is used synonymously.

[154] The average difficulty level is calculated based on the averages per round. The last task bloc of a certain round is excluded from the calculation of the average difficulty level per round in order to avoid distortion because of unfinished task blocs. When only one task bloc was solved in a period (relates to 22 out of 900 observations), the difficulty level of this bloc is included.

[155] Normality tests on performance / attention towards task dimensions: 1) Shapiro-Wilk-Test: z=7.14 / 5.29, p=0.00 / 0.00; 2) Skewness/Kurtosis tests: adj. $\chi^2(2)$=73.47 / 23.76, p=0.00 / 0.00, Skewness=2.48 / 1.16, Kurtosis=12.55 / 4.13.

[156] Cf. Mann and Whitney (1947); Acock (2006), p. 144.

[157] Cf. Brown and Mood (1951); Acock (2006), p. 145; Vorberg and Blankenberger (1999), pp. 158-160

[158] Outliers are defined as observations which are at least 1.5 interquartile ranges (IQRs) below the first quartile or above the third quartile of the distribution, cf. Hoaglin et al. (1986). Extreme outliers are at least 3.0 IQRs below the first quartile or above the third quartile. See Online-Appendix 6 and Online-Appendix 7 for boxplots on the dependent variables performance and difficulty level by treat- ment which illustrate the (extreme) outlying observations.

[159] This corresponds to an approach used by prior experimental studies (e.g. Hannan et al. (2008), Bracha and Fershtman (2013)). However, it is controversial whether individual-level observations within a ranking group are independent from each other.

[160] Order of p-values: Mann-Whitney test / Pearson Chi-Squared test.

[161] Levene's test indicate that there is overall no significant variance heterogeneity between treat- ments: F=0.86, p=0.51 (cf. Levene (1960), results are robust under nonnormality).

[162] Skewness/Kurtosis tests on performance under unidimensional RPI and performance-based contract / flat wage contract: adj. $\chi^2(2)$=22.57 / 3.86, p=0.00 / 0.15, Skewness=2.50 / 0.63. See Online-Appendix 8 and Online-Appendix 9 for histograms on the dependent variables performance and difficulty level by treatment.

[163] The extreme outlying observations relate to 2 participants with total points of 827 (IQR: 3.28) and 1312 (IQR: 6.79).

[164] P-values are not included in Table 1.

Table 1: Dependent Variable Performance – Summary Statistics and Test of Hypotheses and Research Questions

mwt = Mann-Whitney test.
pct = Pearson Chi-squared test on the equality of medians.
P-values compare treatment groups and treatments according to the formulated hypotheses and research questions.
Treatment groups: Unidimensional RPI compared No RPI; Multidimensional RPI compared to Unidimensional RPI.
Treatments: No RPI compared within treatment group; Unidimensional RPI/Perf.-based compared within treatment group to Flat Wage, Unidimensional RPI/Flat Wage compared between treatment groups towards No RPI/Flat Wage.
Multidimensional RPI compared between treatment groups towards Unidimensional RPI.
* Significant at 10%; ** Significant at 5%; *** Significant at 1% level.

RPI		**None**		**Unidimensional**		**Multidimensional**	
Panel A	Mean	285,74		347,74		312,6	
	Median	255		327,5		275,5	
	s.d.	128,3997		209,5149		188,3382	
	mwt	-		0.06* [H1]		0.19 [RQ2a]	
	pct	-		0.05** [H1]		0.11 [RQ2a]	
Compensation Contract		Flat Wage	Perf.-based	Flat Wage	Perf.-based	Flat Wage	Perf.-based
Panel B	Mean	271,72	299,76	332,12	363,36	323,04	302,16
	Median	249	300	331	321	315	234
	s.d.	128,0487	129,8227	164,2331	249,2813	147,8551	224,3486
	Range	144-772	107-620	23-744	87-1312	134-762	88-1217
	n	25	25	25	25	25	25
	mwt	-	0,35	0.05**	0.82 [H2]	0.67 [RQ2b]	0.11 [RQ2b]
	pct	-	0,16	0.01***	0.40 [H2]	0.4 [RQ2b]	0.09* [RQ2b]

Table 2: Dependent Variable Difficulty Level – Summary Statistics and Test of Hypotheses and Research Questions

mwt = Mann-Whitney test.
pct = Pearson Chi-squared test on the equality of medians.

RPI		**None**		**Unidimensional**		**Multidimensional**	
Panel A	Mean	2,77		2,88		2,75	
	Median	2,69		2,51		2,58	
	s.d.	0,97		1,16		1,05	
	mwt	-		0,95		0,72	
	pct	-		0,69		1	
Compensation Contract		Flat Wage	Performance-based	Flat Wage	Performance-based	Flat Wage	Performance-based
Panel B	Mean	2,53448	3	2,58	3,18	2,69	2,81
	Median	2,26	2,75	2,31	2,75	2,64	2,51
	s.d.	0,8371085	1,04	0,98	1,28	1,14	0,98
	Range	1.01 - 4.75	1.61 - 5.67	1.23 - 5.58	1.86 - 6.00	1.13 - 6.00	1.57 - 6.00
	mwt	-	0,11	0,93	0.10*	0,81	0,36
	pct	-	0,16	0,78	0,16	0,4	0,16

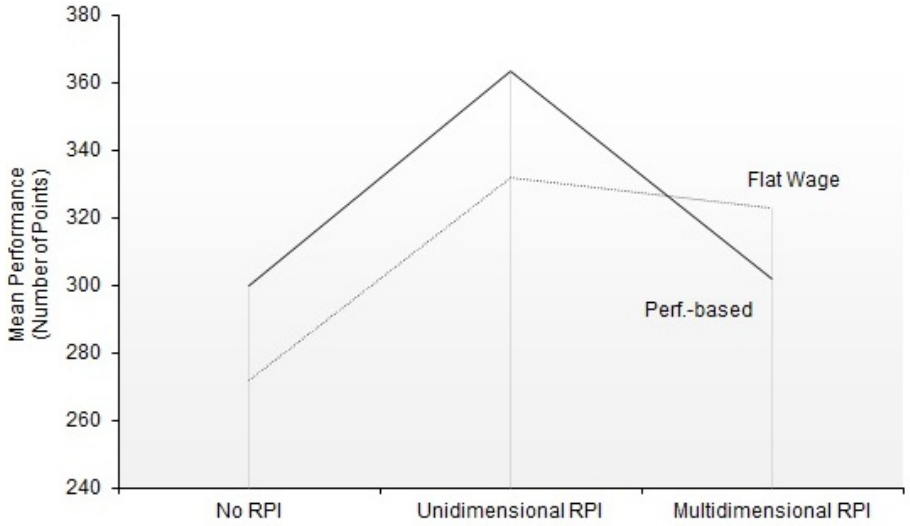

Figure 1: Effect of RPI and Financial Incentives on Mean Performance

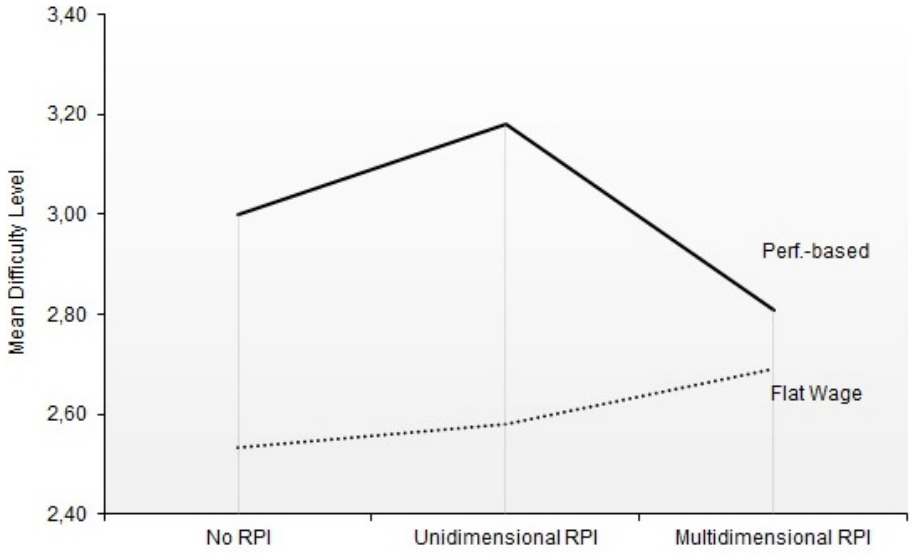

Figure 2: Effect of RPI and Financial Incentives on Mean Difficulty Level

suggests that median performance is significantly different when multidimensional RPI instead of unidimensional RPI is used in conjunction with performance-based compensation (p=0.11 / 0.09).[165] Additionally, it should be noted that the standard deviation of performance is also remarkably higher under a performance-based contract compared to a flat wage contract which is similar to the characteristics of the data on performance under unidimensional RPI.

The examined research questions which are concerned with the effects of the dimensionality of RPI on performance are derived from the research questions RQ1a and RQ1b that deal with possibly varying influences of unidimensional and multidimensional RPI on attention towards task dimensions. Considering RQ1a, the data in Panel A of Table 2 reveal that the mean difficulty level at which participants changed from one bloc to another is lower under multidimensional compared to unidimensional RPI. However, it is also shown that the median difficulty level is slightly higher under multidimensional RPI. Overall, there is no significant difference in attention towards task dimensions between the two forms of RPI (p=0.72 / 1.00). Also, a comparison between the treatments that involve the provision of multidimensional RPI and the 'No RPI' treatments does not imply any significant variation (p=0.69 / 0.69)[166].

The puzzling results derived from the pooled data are driven by differences in the impact of the two forms of RPI

[165] The Mann-Whitney test do not indicate significance because of an upward outlier under multidimensional RPI and performance-based compensation. The related participant reached a total number of points of 1217 (IQR: 6.38).

[166] P-values are not included in Table 2.

on attention towards task dimensions under performance-based and flat wage compensation. As reported in Panel B of Table 2, the application of a performance-based contract is associated with an increase in mean and median difficulty level when participants receive unidimensional RPI or no RPI. The Mann-Whitney test indicates that this difference in attention allocation between the compensation contracts is significant for unidimensional RPI (p=0.10 / 0.16).[167] With regard to multidimensional RPI, the mean difficulty levels suggest a slight increase when performance-based compensation is applied but the median values show a reduction. The indication of a positive effect is caused by an extreme upward outlier[168] under performance-based pay and differences in the skewness of the distributions.[169] A comparison between the effects of the two forms of RPI under different financial incentives reveal that multidimensional compared to unidimensional RPI induces a higher difficulty level under a flat wage compensation but a lower difficulty level in the presence of a performance-based contract. Figure 2 graphically summarizes the described interaction between the dimensionality of RPI and the used compensation contract. However, concerning RQ1b, no significant difference can be found in attention towards task dimension when participants receive multidimensional compared to unidimensional RPI under performance-based compensation (p=0.36 / 0.16). Likewise, there is no significant difference when a flat wage contract is applied (p=0.81 / 0.40).

So far, it can be summarized that unidimensional RPI significantly increases participant's performance in the examined multidimensional task setting even though this effect is not magnified under performance-based compensation which contradicts the formulated prediction. Importantly, the results of the bivariate analysis indicate that multidimensional RPI has no significant positive effect on performance but that it rather induces a performance decrease compared to unidimensional RPI in the presence of a performance-based contract. The negative performance effect is accompanied by a lower difficulty level at which participants changed from one task bloc another. However, this difference in attention towards task dimensions between the two forms of RPI under performance-based compensation appears not to be significant. Overall, it can be also noted that both the provision of unidimensional and multidimensional RPI results in a considerably higher spread of performance as compared to when participants receive no RPI.

[167]The Pearson Chi-Squared test does not indicate significance since the difference is strongly driven by two observations which relate to a participants with an average difficulty level of 5.7 and 6.0.

[168]The related participant reached an average difficulty level of 6 (IQR: 6.81).

[169]Skewness/Kurtosis tests on attention towards task dimensions under multidimensional RPI and performance-based contract / flat wage contract: adj. $\chi^2(2)$=7.42 / 13.48, p=0.03 / 0.00, Skew- ness=1.66 / 1.08. See Online-Appendix 8 and Online-Appendix 9 for histograms on the dependent variables performance and difficulty level by treatment.

5.1.2. Multivariate analyses

Based on the results of the bivariate analyses, this section aims to provide a more differentiated understanding of the effects of the dimensionality of RPI in the presence of financial incentives. For this purpose, multiple linear regressions are conducted which control for further determinants of the dependent variables and consider interaction effects. In particular, three models are specified which are tested with an OLS regression as well as a quantile regression approach. The latter is defined for the 50th percentile and involves median regressions. Notably, it is included in the analysis because of its non-parametric character and its robustness to outlying observations.[170]

In Table 3, results are reported for the dependent variable performance. The findings of the median regression analysis are only shown for model (3) since its results for model (1) and model (2) do not considerably differ from the presented OLS estimations.[171] In order to ensure the robustness of the OLS regression results, heteroskedasticity-consistent standard error estimators are implemented.[172] In this context, it should be also noted that post-estimation analyses indicate no threat of multicollinearity to the results.[173]

Model (1) consists of dummy variables indicating the provision of unidimensional RPI, multidimensional RPI as well as performance-based compensation. The results imply that performance is significantly higher under unidimensional RPI (p=0.077) while the positive performance difference between multidimensional RPI and no RPI is not statistically significant (p=0.408). Also, the influence of performance-based compensation appears to be positive but the coefficient is insignificant (p=0.663). However, overall, the explanatory power of model (1) is low and the null hypothesis cannot be rejected that all of the model coefficients are zero (Prob>F=0.354, R^2=0.021).

In model (2), personal variables are added so as to control for differences in participants' characteristics between treatments. Specifically, the participants' age and risk aversion are included as well dummy variables that divide for gender and indicate whether participants major in a math-related field of study. Additionally, model (2) controls for participants' preferences for competition because prior literature suggest them to be determinant of performance in competitive task settings.[174] Importantly, the results reveal that the effects of the dimensionality of RPI and of the applied compensa-

[170]Cf. Koenker and Bassett (1978); Koenker (2005).

[171]A comprehensive overview of the results of the median regressions for model (1) - (3) can be found in Online-Appendix 10.

[172]Breusch-Pagan test (cf. Breusch and Pagan (1979)) and White's test (cf. White (1980)) are applied in order to check for the homoscedasticity of residuals. The Breusch-Pagan test (normality assump- tion removed) indicates heteroskedasticity for model (2) and model (3) ($\chi^2(1)$=5.51, p=0.019; $\chi^2(1)$=8.87, p=0.003).

[173]Estimated variance inflation factors (VIF) are constantly below the value 10 and thus indicate that there is no problematic degree of multicollinearity (cf. Marquaridt (1970)). VIF ranges are as follows: model (1): 1.00-1.33; model (2): 1.02-1.43; model (3): VIF=1.11-3.39.

[174]E.g. prior studies are concerned with gender differences in preferences

Table 3: Dependent Variable Difficulty Level – Regression Results

Model (1) reports treatment group effects only. Model (2) additionally includes the control variables Gender (dummy variable, 1='male', 2='female'), Age (mean-centered), Math-related Field of Study (dummy variable, 1='Math-related Field of Study', 2='No Math-related Field of Study), Risk Aversion (range 0 - 10 where 0 indicates 'highly risk loving' and 10 indicates 'highly risk averse'), Preference for Competition (participants' responses on a seven-point Likert scale to the statement 'I perform better when I compete against someone rather than when I am the only one striving for a goal', 1='not at all', 7='to a great degree'). Model (3) adds interaction terms between the treatment groups for RPI and for compensation scheme as well as interaction terms between the treatment groups for RPI and a dummy variable for participants' position in the final ranking on overall performance (1='Rank 4 or 5', 2='Rank 1, 2 or 3'). Heteroskedasticity-consistent standard error estimators are implemented in the OLS regression.
* Significant at 10%; ** Significant at 5%; *** Significant at 1% level.

Regressors (DV: Performance)	**Model (1) OLS Reg.**	**Model (2) OLS Reg.**	**Model (3) OLS Reg.**	**Model (4) OLS Reg.**
	Coeff. (S.E.) p-value	Coeff. (S.E.) p-value	Coeff. (S.E.) p-value	Coeff. (S.E.) p-value
Unidimensional RPI	62 (34.782) 0.077*	71.173 (39.136) 0.071*	159.7 (45.774) 0.001***	149.72 (27.287) 0.000***
Multidimensional RPI	26.86 (32.358) 0.4078	14.199 (32.143) 0.6594	128.8 (41.38) 0.002***	99.212 (26.656) 0.000***
Perf.-based Compensation	12.8 (29.27) 0.6625	20.139 (27.806) 0.4701	43.008 (35.829) 0.2321	74.395 (24.833) 0.003***
Gender		2.7174 (27.933) 0.9226	10.682 (23.853) 0.655	-6.598 (15.269) 0.6664
Age		10.473 (4.0629) 0.011**	9.3839 (3.8275) 0.015**	6.3 (2.0672) 0.003***
Math-related Field of Study		63.401 (40.659) 0.1212	24.004 (34.228) 0.4843	45.385 (19.366) 0.021**
Risk aversion		-14.34 (10.517) 0.1748	-15.73 (8.587) 0.069*	-10.25 (4.3635) 0.020**
Preference for Competition		15.696 (7.7417) 0.044**	5.6771 (7.0996) 0.4253	2.1902 (3.7655) 0.5618
Unidimensional RPI x Perf.-based Compensation			-12.95 (56.512) 0.819	-69.86 (35.073) 0.048**

(Continued)

for competition that impact performance outcomes, cf. Croson and Gneezy (2009); Niederle and Vesterlund (2007).

Table 3—continued

Regressors (DV: Performance)	**Model (1) OLS Reg.**	**Model (2) OLS Reg.**	**Model (3) OLS Reg.**	**Model (4) OLS Reg.**
Multidimensional RPI x Perf.-based Compensation			-67.1	-91.62
			(56.761)	(35.112)
			0.2392	0.010***
Unidimensional RPI x Low Rank			-208.5	-132.3
			(46.048)	(25.535)
			0.000***	0.000***
Multidimensional RPI x Low Rank			-197.2	-169
			(41.653)	(26.248)
			0.000***	0.000***
Constant	279.34	245.85	310.64	253.21
	(23.245)	(78.173)	(72.296)	(34.588)
	0.000***	0.002***	0.000***	0.000***
N	150	150	150	150
Prob>F	0.3538	0.0098	0	-
R^2	0.0214	0.1108	0.3025	-
Pseudo R^2	-	-	-	0.2145

tion contract remain largely unchanged when personal variables are considered. Yet, it is also shown that participants' age as well as their preference for competition both significantly positively affect performance (p=0.011, p=0.044). Furthermore, model (2) in general is statistical significant and has thus a higher explanatory power than model (1) (Prob>F=0.010, R^2=0.111).

Model (3) introduces interaction terms which account for differences in the performance effects of the dimensionality of RPI under performance-based and flat wage compensation. Also, it contains interaction variables that distinguish between the influences of the two forms of RPI on high ranked and low ranked participants.[175] In this way, the model addresses findings from the bivariate analyses which indicate that performance effects of RPI differ with regard to the applied compensation contract as well as that the standard deviation of performance is comparably higher in the presence of RPI.

Overall, the inclusion of the interaction terms substantially increases the predictive power of the model (Prob>F =0.000, R^2=0.303). The results are reported for an OLS as well as a median regression. Importantly, both statistical analyses indicate that performance significantly rises under unidimensional RPI (p=0.001 / 0.000)[176] as well as under multidimensional RPI (p=0.002 / 0.000). The median regression finds moreover a significant positive effect of performance-based compensation (p=0.003) while the related OLS estimates are still insignificant (p=0.232). By contrast, however, the interactions between performance-based compensation and unidimensional as well as multidimensional RPI are both negative (p=0.819 / 0.048; p=0.239 / 0.010)[177]. The results are again significant for the median regression but not for the OLS regression which may be explained by the already mentioned extreme upward outliers that are present in the treatments that involve RPI and performance-based compensation. Notably, the size of the coefficient of the interaction term is larger for multidimensional RPI than for unidimensional RPI. In this context, the significant negative interaction between multidimensional RPI and performance-based pay in the median regression has almost the same size as the positive performance effect indicated by the dummy variable for multidimensional RPI. Therefore, the overall effect of the application of multidimensional RPI under performance-based compensation is nearly zero. As opposed to that, OLS and median regression results both imply that unidimensional RPI overall impacts performance considerably positively when performance-based compensation is applied.

Further on, model (3) shows significant negative interactions between the two forms of RPI and the overall rank-

[175] High ranked participants are characterized by a final rank of 1, 2 or 3 on overall performance. Low ranked participants are characterized by a final rank of 4 or 5 on overall performance.

[176] Order of p-values: OLS regression / Median regression.

[177] The p-values that refer to the effect of unidimensional RPI are mentioned first.

ing position of the participants (p=0.000 / 0.000; p=0.000 / 0.000). Specifically, in the case of multidimensional RPI, the size of the coefficients of the OLS as well as the median regression indicate that the performance of low ranked participants overall decreases as compared to when they receive no RPI. For unidimensional RPI, the same can be concluded with regard to the estimates of the OLS regression. Also, the negative interaction effect appears to be greater for unidimensional than for multidimensional RPI considering the OLS results. Yet, the opposite holds for the findings of the median regression. Additional analyses disclose that these discrepancies between the statistical analyses are caused by a downward outlier[178] included in the observations on performance under unidimensional RPI which relates to a low ranked participant. Lastly, with respect to the influence of personal variables on performance, model (3) confirms the positive influence of participants' age (p=0.015 / 0.003) but also predicts a significant negative effect of risk aversion (p=0.069 / 0.020). Furthermore, the positive impact of participants' preference for competition is no longer significant (p=0.425 / 0.562) while the median regression indicates that majoring in a math-related field of study significantly increases performance (p=0.484 / 0.021).

Overall, post-estimation analyses applied to the presented models reveal that the performance effects of unidimensional and multidimensional RPI under flat wage compensation do not significantly differ from each other.[179] Similarly, in model (3), no significant differences are found between the interaction effects of the two forms of RPI with the applied compensation scheme[180] and the overall ranking position of the participants[181].

Table 4 reports regression results for the dependent variable attention towards task dimensions. The presented models correspond to the models applied in the regression analysis of the dependent variable performance. The estimates of the median regression are not included in Table 4 since they do not involve additional insights.[182] Like in the prior analysis on performance, heteroskedasticity-consistent standard error estimators are used with regard to the robustness of the OLS regression results.[183] Also, applied post-estimation analyses show no threat of multicollinearity.[184]

To begin with model (1), unidimensional RPI seem to increase the difficulty level at which participants change from one bloc to another while the effect of multidimensional RPI appears to be negative as compared to the treatments that do not involve RPI. However, neither of the two forms of RPI significantly influence attention towards task dimensions (p=0.603, p=0.921). By contrast, the results show that the application of performance-based compensation is associated with a significant increase in the average difficulty level (p=0.023). Yet, the explanatory power of model (1) is relatively low and the model coefficients are overall not significantly different from zero (Prob>F=0.146, R^2=0.038).

In model (2), the addition of personal variables reveals that participants' risk aversion significantly lowers the difficulty level at which they change from one task bloc to another (p=0.001). On the other hand, it also indicates a significant positive effect of participants' preference for competition and of majoring in a math-related field of study (p=0.045, p=0.040). Apart from that, the effects of the two forms of RPI and of the compensation scheme do not considerably change as compared to model (1). In general, model (2) has a higher explanatory power than model (1) and is statistically significant (Prob>F=0.001, R^2=0.174).

Next, model (3) includes interaction terms between the dimensionality of RPI and the compensation scheme as well as the total rank of participants. Therewith, the predictive power of the model is increased (Prob>F=0.001, R^2=0.209). The results imply positive but insignificant coefficients for the effects of unidimensional and multidimensional RPI on the average difficulty level (p=0.113, p=0.329). Similar to model (1) and (2), the OLS estimates demonstrate a significant increase of the difficulty level under performance-based compensation (p=0.048). There seems to be no interaction between performance-based compensation and the provision of unidimensional RPI since the related coefficient is nearly zero (p=0.985). As opposed to that, the analysis reveals that multidimensional RPI overall decreases the average difficulty level under performance-based compensation in comparison to when participants receive no RPI. Although the coefficient is not significant, its size suggest that the interaction effect outweighs the positive influence indicated by the dummy variable for multidimensional RPI (p=0.246).

Model (3) additionally discloses that the interaction effects between the two forms of RPI and the overall ranking position of the participants are both negative. Yet, the negative interaction is only significant in case of unidimensional RPI (p=0.040, p=0.448). Notably, the coefficients show that the application of unidimensional RPI overall lowers the average difficulty level of low ranked participants as compared to the treatments which do not involve RPI. By contrast, the overall effect of multidimensional RPI on the average difficulty level remains to be positive with regard to low ranked participants. Finally, in model (3), the risk aversion of participants still significantly reduces the average difficulty level at which they change from one bloc to another (p=0.001). Besides, no other personal variable has a considerable influence on attention towards task dimensions.

[178] The related participant only reached a total number of points of 23 (IQR: - 1.80).

[179] Wald tests are applied in order to test for the equality of the coefficients (cf. Wald (1943)). Model (1): F=0.77, p=0.381; Model (2): F=1.77, p=0.186; Model (3): F=0.40, p=0.526.

[180] Wald test: F=0.72, p=0.398.

[181] Wald test: F=0.04, p=0.846.

[182] A comprehensive overview of the results of the median regressions for model (1) - (3) can be found in Online-Appendix 11.

[183] Breusch-Pagan test (cf. Breusch and Pagan (1979)) and White's test (cf. White (1980)) are conducted to check for the homoscedasticity of residuals. The Breusch-Pagan test (normality assumption removed) indicates heteroskedasticity for model (2) and model (3) ($\chi^2(1)$=6.09, p=0.014; $\chi^2(1)$=6.39, p=0.012).

[184] Estimated VIFs are constantly below the value 10 (cf. Marquaridt (1970)). VIF ranges are as follows: model (1): 1.00-1.33; model (2): 1.02-1.43; model (3): VIF=1.11-3.39.

Table 4: Dependent Variable Difficulty Level – Regression Results

Model (1) reports treatment group effects only. Model (2) additionally includes the control variables Gender (dummy variable, 1='male', 2='female'), Age (mean-centered), Math-related Field of Study (dummy variable, 1='Math-related Field of Study', 2='No Math-related Field of Study), Risk Aversion (range 0 - 10 where 0 indicates 'highly risk loving' and 10 indicates 'highly risk averse'), Preference for Competition (participants' responses on a seven-point Likert scale to the statement 'I perform better when I compete against someone rather than when I am the only one striving for a goal', 1='not at all', 7='to a great degree'). Model (3) adds interaction terms between the treatment groups for RPI and for compensation scheme as well as interaction terms between the treatment groups for RPI and a dummy variable for participants' position in the final ranking on overall performance (1='Rank 4 or 5', 2='Rank 1, 2 or 3'). Heteroskedasticity-consistent standard error estimators are implemented.
* Significant at 10%; ** Significant at 5%; *** Significant at 1% level.

Regressors (DV: Performance)	**Model (1) OLS Reg.**	**Model (2) OLS Reg.**	**Model (3) OLS Reg.**	**Model (4) OLS Reg.**
	Coeff. (S.E.) p-value	Coeff. (S.E.) p-value	Coeff. (S.E.) p-value	Coeff. (S.E.) p-value
Unidimensional RPI	0.1087 (0.2083) 0.6027	0.2514 (0.2052) 0.2224	0.4612 (0.2891) 0.113	
Multidimensional RPI	-0.02 (0.2009) 0.9214	-0.017 (0.2027) 0.9325	0.3228 (0.3298) 0.3294	
Perf.-based Compensation	0.394 (0.1713) 0.023**	0.4078 (0.1654) 0.015**	0.5566 (0.2795) 0.048**	
Gender		0.2393 (0.1689) 0.1588	0.2847 (0.1736) 0.1034	
Age		0.0349 (0.025) 0.1659	0.0337 (0.0246) 0.1727	
Math-related Field of Study		0.3977 (0.1915) 0.040**	0.3096 (0.2118) 0.1461	
Risk aversion		-0.176 (0.051) 0.001***	-0.174 (0.049) 0.001***	
Preference for Competition		0.0805 (0.0399) 0.045**	0.0648 (0.0409) 0.1151	
Unidimensional RPI x Perf.-based Compensation			0.0075 (0.397) 0.9849	

(Continued)

Table 4—continued

Regressors (DV: Performance)	**Model (1) OLS Reg.**	**Model (2) OLS Reg.**	**Model (3) OLS Reg.**	**Model (4) OLS Reg.**
Multidimensional RPI x Perf.-based Compensation			-0.47	
			(0.4033)	
			0.2461	
Unidimensional RPI x Low Rank			-0.553	
			(0.2669)	
			0.040**	
Multidimensional RPI x Low Rank			-0.251	
			(0.3295)	
			0.4477	
Constant	2.5726	2.7851	2.8106	
	(0.1499)	(0.4243)	(0.4456)	
	0.000***	0.000***	0.000***	
N	150	150	150	
Prob>F	0.1464	0.0006	0.0005	
R^2	0.0377	0.1737	0.2093	

Post-estimation analyses are applied in order to test whether the effects of unidimensional and multidimensional RPI significantly differ from each other in the presented models. Overall, no significant variation is found between the influences of the two forms of RPI on attention towards task dimensions under flat wage compensation.[185] Also, with regard to model (3), the interactions of unidimensional and multidimensional RPI with the compensation scheme[186] as well as with the overall ranking position of the participants[187] do not show significant differences.

5.1.3. Discussion

The effects of unidimensional RPI on performance and attention towards task dimensions (H1, H2)

The basis for the examination of the influence of the dimensionality of RPI on performance and attention towards task dimensions is laid by the first hypothesis H1 stating that unidimensional RPI increases performance in multidimensional task settings. The results of both the bivariate and the multivariate analyses fully support the formulated prediction. This substantiates the expected positive effects of RPI on an individual's motivation and learning which imply higher effort and the alignment of attention allocation towards overall task performance. Therefore, the findings of prior research on performance improvements under RPI appear to be generalizable to multidimensional task settings when RPI compares peer performance based on an overall performance measure that aggregates relevant task dimensions.

However, regarding the second hypothesis H2, the analysis does not confirm the expected performance increase when unidimensional RPI is provided together with performance-based instead of flat wage compensation. Particularly, neither an additive effect nor a positive interaction effect can be identified with respect to the motivational impact of performance-based compensation predicted by agency theory.[188] By contrast, results of the median regression suggest that unidimensional RPI significantly negatively interacts with the application of performance-based pay while its overall performance effect still stays considerably positively. With regard to the reasons of this discrepancy between the theoretical prediction and the experimental results, it is firstly of interest if there are unexpected differences in participants' attention towards task dimensions when unidimensional RPI is provided together with performance-based instead of flat wage compensation. Indeed, the bivariate analysis indicates a significant increase in the mean difficulty level under a performance-based contract. However, the corresponding interaction term in the regression analysis on attention towards task dimensions is nearly zero and insignificant. In this context, additional tests reveal that the result of the bivariate analysis is strongly driven by two observations in the data set

[185] Cf. Wald (1943); Wald test: Model (1): F=0.34, p=0.560; Model (2): F=1.70, p=0.194; Model (3): F=0.15, p=0.695.

[186] Wald test: F=1.36, p=0.245.

[187] Wald test: F=0.52, p=0.473.

[188] The graphical comparison of mean performance values in Figure 1 in section 5.1.1 originally suggested an additive effect. However, this effect on mean performance did not hold under an OLS regression which included control variables. The median performance values already indicated in 5.1.1 a decrease of performance which was confirmed in the median regression.

which relate to participants with an average difficulty level between 5.5 and 6.0.[189] Notably, the same participants also achieved above-average performance[190] and further analyses show that an increase in the average difficulty level under performance-based compensation can be mostly attributed to high performing participants.[191] Therefore, it appears that the introduction of performance-based compensation is not associated with an undesirable distortion in attention allocation under unidimensional RPI which might have impede performance improvements.

Importantly, the distribution of performance under unidimensional RPI suggests that the application of a performance-based contract has actually resulted in a positive effect with regard to high performing participants.[192] Also, the presence of extreme upward outlying observations under the performance-based scheme may be an indication that the predicted motivational effect of financial incentives has pushed participants with high ability to achieve maximum performance.[193] With respect to the participants with below-average performance, it should be emphasized that the nature of the applied task implies that the relationship between effort and performance is moderated by participants' cognitive abilities. Particularly, performance improvements through higher motivation under performance-based compensation are limited based on participants' ability to solve multiplication problems.[194] In this regard, performance-based compensation may have only positively affected performance of participants that did not receive RPI since the provision of RPI already improved performance under a flat wage contract which is why there might have been no much more room for improvement.

Additionally, it is worth mentioning that the application of a variable payment increases the performance pressure inherent in the setting which might be critical in conjunction with the competitive pressure induced by RPI. Notably, it could have resulted in a so called 'choking under pressure' effect which refers to the occurrence of inferior performance under pressure circumstances.[195] In this context, Beilock et al. (2004) demonstrate that pressure induced by monetary rewards harms performance in mathematical problem solving which requires heavy working memory such as multiplication tasks.[196] In the experimental setting, especially participants with below-average performance could have been affected by 'choking under pressure' since perceived pressure might increase through a low ranking position as well as the threat that monetary earnings are falling short of expectations.[197] However, additional analysis on the participants' mistake share and on the average time required to solve a problem of a certain difficulty level do not further underpin this supposition.[198]

Differences between the effects of unidimensional and multidimensional RPI on performance and attention towards task dimensions (RQ1a, RQ1b, RQ2a, RQ2b)

Besides the formulated hypotheses, the study has developed two-tailed research questions so as to address its objective of comparing the performance and attention allocation effects of unidimensional RPI with that of multidimensional RPI. Research question RQ2a is concerned with performance differences. Firstly, the bivariate as well as the multivariate analyses do not find a significant variation between the performance effects of the two forms of RPI. Yet, there is also no evidence for a significant positive performance effect of multidimensional RPI unless the regression analysis considers its interaction effects with the applied compensation scheme and the overall ranking position of the participants. Similar to unidimensional RPI, the results of the median regression reveal a significant negative interaction between multidimensional RPI and performance-based pay.[199] Importantly, however, the overall effect of multidimensional RPI under performance-based compensation is therefore nearly zero while unidimensional RPI still induces a remarkable performance increase. Regarding research question RQ2b, this complies with the finding of the bivariate analyses that median performance significantly differs between the two forms of RPI in the presence of a performance-based contract.

In order to interpret the described performance effects, potential differences in attention towards task dimensions between unidimensional and multidimensional RPI have to be considered. In relation to research question RQ1a, the analyses overall indicate that the average difficulty level under the two forms of RPI is not significantly different which

[189] When the observations are excluded from the data set, neither the Mann-Whitney test nor the Pearson Chi-Squared test indicate significant differences between the average difficulty level under perf.-based and flat wage compensation (p=0.22 / 0.39)

[190] The related participants achieved a total number of points of 602 (difficulty level 6.0) and 396 (difficulty level 5.7).

[191] When the sample is restricted to participants that perform above the mean of their treatment, mean /median difficulty levels are the following: Perf.-based: 3.82 / 3.56; Flat wage: 2.74 / 2.49. The difference is not significant (p=0.18 / 0.52).

[192] When the sample is restricted to participants that perform above the mean of their treatment, mean /median performance is the following: Perf.-based: 629.29 / 492; Flat wage: 453.42 / 389. The difference is not significant (p=0.12 /0.11).

[193] The extreme outlying observations relate to 2 participants with total points of 827 (IQR: 3.28) and 1312 (IQR: 6.79).

[194] Prior literature commonly argues that the positive performance effect of financial incentives predicted by agency theory is limited when tasks involve cognitive performance, cf. Bonner and Sprinkle (2002), p. 320; Camerer and Hogarth (1999); Rydval and Ortmann (2004).

[195] Cf. Baumeister (1984), p. 610; Ariely et al. (2009).

[196] Beilock et al. (2004) relates this finding to distraction theories which propose that pressure increases thoughts about the situation and its importance which reduce the capacity of the working memory that is available for the execution of the task.

[197] Commonly, participants in the experimental laboratory MELESSA expect to earn between 10 and 15 Euro per hour of participation. Since the show-up fee accounts for 4 Euro, participants had to achieve 300 points in order to assure that they earn at least 10 Euro in the experiment. (The earn- ings from the risk game are not disclosed until the end of the experiment and could have been in the minimum only 0.1 Euro.)

[198] Mistake share: (Number of incorrect answers)/ (Sum of number of correct and incorrect answers per difficulty level).

[199] As shown in section 5.1.2, the interaction effects are not significantly different from each other.

corresponds to the results on the performance effects. Further on, the introduction of performance-based compensation seem to differently affect the average difficulty level reached in the presence of unidimensional and multidimensional RPI. More concretely, the regression analysis reveals a negative interaction effect between multidimensional RPI and performance-based compensation and implies an overall decrease of the average difficulty level in comparison to the treatment that do not involve RPI. This contrasts with the very small and positive interaction between unidimensional RPI and performance-based pay. However, with respect to RQ1b, the effects of multidimensional RPI on attention towards task dimension are not statistically significant and neither is the difference between the interaction terms of the two forms of RPI with performance-based compensation.

From a theoretical perspective, it was outlined that multidimensional as compared to unidimensional RPI may differently affect attention allocation and performance because it do not limit peer comparisons to the overall performance measure which can promote learning but also motivate a distortion of attention. In the presence of performance-based compensation, the rewarding of overall performance were expected to weaken potential attention distortion effects and to strengthen the impact of learning benefits. Therefore, it surprises that negative performance effects of multidimensional RPI are present under performance-based compensation but not when a flat wage is applied. More in-depth analyses reveal that the decrease in performance between the two compensation contracts do not hold with regard to high performing participants which complies with the findings on unidimensional RPI.[200] Also, it can be shown that the increase in the average difficulty level suggested by the mean values, which opposes the indication of the median values and the regression results, can be attributed to those high performing participants.[201] In this context, a potential attention distortion towards the 'quantity' dimension of the task for participants performing below the treatment mean is further investigated. Notably, with respect to participants that achieved rank one or two in the 'quantity' ranking of their group, overall performance is considerably lower for multidimensional RPI under performance-based as compared to flat wage compensation. Also, the performance under both compensation contracts falls short to what is achieved by the top two performers of the 'quantity' ranking under unidimensional RPI.[202] This may indicate that participants provided with multidimensional RPI concentrate on the 'quantity' dimension of the task at the expense of overall performance, especially under performance-based compensation. However, the differences between the treatments are not significant and it appears that the performance decrease under multidimensional RPI in the presence of a performance-based contract might be also strengthened by further determinants.

As already mentioned before, the performance pressure inherent to a task setting is increased by the provision of RPI as well as by the application of a performance-based contract. Therefore, performance deteriorations of participants that are exposed to both forms of pressure might be linked to a 'choking under pressure' effect. In contrast to unidimensional RPI, additional analyses on the average time that participants under multidimensional RPI have required to solve a problem of a certain difficulty level provide some support for this supposition. Specifically, it can be shown that participants provided with performance-based compensation have spent more time on problems of all difficulty levels in comparison to those that received a flat wage. Notably, the differences between the compensation contracts are significant for difficulty level 1 and 2 when outliers are excluded from the sample.[203] Based on this finding, it may be argued that the provision of performance-based compensation increases pressure more under multidimensional RPI than under unidimensional RPI. From a theoretical perspective, this could be explained by the additional cognitive effort that is required when individuals have to solve trade-offs between the potentially conflicting motives induced by in the rewarding of overall performance and the facilitation of competition on dimension-specific performance measures.[204]

The effects of unidimensional and multidimensional RPI on performance and attention towards task dimensions of low ranked participants

Lastly, with regard to the overall performance effects of unidimensional and multidimensional RPI, the regression analysis reveals significant negative interactions between both forms of RPI and the overall ranking position of the participants.[205] In particular, the size of the effects of the OLS regression indi-

[200] When the sample is restricted to participants that perform above the mean of their treatment, mean / median performance is the following: Perf.-based: 517.63 / 447.5; Flat: 455.9 / 420.5. The difference is not significant (p=0.93 /1.00). Similarly to unidimensional RPI, the sample on perf.- based compensation also includes an extreme outlying observation. The related participant achieved a total number of points of 1217 (IQR: 6.38).

[201] When the sample is restricted to participants that perform above the mean of their treatment, mean /median difficulty levels are the following: Perf.-based: 3.23 / 2.76; Flat: 3.02 / 2.64. The dif- ference is not significant (p=0.96 / 1.00). The distribution under perf.-based compensation includes two outlying observation related to participants with an average difficulty level of 4.71 (IQR: 2.16) and 6.0 (IQR: 3.81). The same participants also represent the two outlying observa- tions on performance. However, performance as well as the average difficulty are for both partici- pants that perform above and below the mean still considerably higher under unidimensional RPI and perf.-based compensation.

[202] Mean and median performance is as follows: Multidimensional RPI: 1) Flat: 322.7 / 327.5, Perf.- based: 296.9 / 303.5; Unidimensional RPI: 1) Flat: 363 / 354, Perf.-based: 354 / 311. The differences are not significant (all p>0.37).

[203] Mean and median time values: Difficulty level 1) Flat: 9.42 / 8.23; Perf.-based: 11.36 / 10.48 (p=0.24 / 0.04); Difficulty level 2) Flat: 10.59 / 10.25; Perf.-based: 13.64 / 12.57 (p=0.04 / 0.14). The differences in the time values are not expected to be driven by different abilities since the treatments do not differ in their share of students majoring in a math-related field of study and since the first two difficulty levels only require basic math skills. There are no significant differ- ences in the average mistake shares per difficulty level.

[204] Cf. Payne et al. (1993), p. 30; Hogarth (1987).

[205] As shown in section 5.1.2, the interaction effects are not significantly different from each other. It should be emphasized that low ranked participants are not necessarily below-average performers of their treatment since the participants compete in groups of five while each treatment is applied to

cate that the performance of low ranked participants overall decreases when they are provided with either form of RPI. In the median regression, this only applies to multidimensional RPI.[206] In interpreting these findings, it should be referred to the insight of social comparison theory that comparisons with superior others can be on the one the hand inspiring and spur the motivation for performance improvement but may also adversely affect self-evaluation and promote negative feelings of inferiority. As outlined earlier in section 2.1.1, the latter is expected to occur when a comparison implies a contrastive evaluation of the comparison target which emphasizes the individual's separation toward it.[207] More concretely, comparison standards are perceived to be not attainable which might have applied to the participants of the experiment which experienced a ranking position below the top three ranks.

In the described context, peer comparisons can firstly decrease task motivation and provoke defensive reactions which can result in lower effort and deteriorating performance.[208] Secondly, threatening upward comparisons can lower self-efficacy beliefs and therefore cause a devotion of cognitive effort towards activities that are unrelated to task execution such as worrying about the own performance and interpreting the ranking outcome.[209] Here, prior experimental studies show that the behavioural responses can also include inconsistent changes of task strategies and the adoption of risky strategies.[210] With regard to the performance decrease of the low ranked participants in the conducted experiment, some evidence is found with regard to the second explanation. Particularly, the responses to the post-experimental questionnaire indicate that low ranked as compared to high ranked participants were more nervous about their rank and perceived that thinking about their rank interfered to a greater extent with their performance (p=0.02, p=0.08).[211] Also, it appears that low ranked participants changed their task strategy more often during the experiment since the number of changes in the average difficulty level between the rounds is significantly greater in comparison with high ranked participants (p=0.01).[212] Moreover, it can be shown that the share of incorrect answers on the total number of answers is significantly higher for low ranked participants (p=0.00).[213] In this context, they have also required significantly more time to solve problems of difficulty level one to five (all p<0.00).[214] However, this can be as well an indication for decreases in the effort level due to a loss of motivation.

As mentioned before, the regression results indicate that the negative performance effect of the provision of RPI on low ranked participants appears to be specifically emphasized under multidimensional RPI. This result may be interpreted in two different ways. First, participants that are provided with multidimensional RPI have the opportunity to reduce the self-threating effect induced by a negative social comparison on overall performance through the affirmation of their competencies in one of the task dimensions. Therefore, the predictions derived from self-affirmation theory with regard to potential attention distortion effects under multidimensional RPI might be especially relevant for low ranked participants.[215] In this regard, the interaction term included in the regression analysis on attention towards task dimensions indicates a decrease in the average difficulty level of low ranked participants. However, the effect is not significant and additional analyses show that there are is also no difference between the two forms of RPI regarding the number of low ranked participants that achieved a high rank in the 'quantity' or the 'value' dimension of the task.[216] Notably, the regression results disclose a significant negative interaction between unidimensional RPI and the average difficulty level of low ranked participants. This might be explained by decreased motivation and a lower effort level since tasks of lower difficulty require less cognitive effort and the provision of unidimensional RPI does not involve potential incentives for a distortion of attention towards one task dimension. Indeed, it can be demonstrated that the significant effect is driven by two participants with a considerably low respectively downward outlying overall performance.[217]

25 participants in total. Therefore, the study compares the overall effect of the two forms of RPI with regard to low ranked participants and disentangles effects between above and below-average performers only in the context of the application of different compensation schemes.

[206] As mentioned before, the data on unidimensional RPI include a downward outlying observation which relates to a low ranked participant (performance: 23 points, IQR: -1.80).

[207] Cf. Festinger (1954); Suls et al. (2002); Corcoran et al. (2011); Tesser (1988).

[208] Cf. Corcoran et al. (2011); Stapel and Schwinghammer (2004). Defensive reactions may include e.g. that individuals question the relevance of the performance dimension, search for differences be- tween themselves and the comparison target or try to self-affirm their competencies in another area.

[209] Cf. Bandura (1977b); Bandura/Wood (1989).

[210] Cf. Hannan et al. (2008); Camerer and Hogarth (1999).

[211] Participants had to indicate rank nervousness and interference using a seven-point Likert scale (1=not at all, 7=to a great degree). The questions are adapted from Tafkov (2013). See Online-Appendix 5 for the questions of the post-experimental questionnaire. Based on the answers, continuous varia- bles are created. Mean values: Rank nervousness - Low rank: 3.18 - High rank: 2.35; Rank inter- ference - Low rank: 2.83 - High rank: 2.15. P-values are derived from a Mann-Whitney test. The comparison is limited to participants that received RPI (unidimensional or multidimensional RPI). There are no significant differences in the responses of low and high ranked participants that were not provided with RPI (and therefore did not know their ranking position).

[212] Mean number of changes in the average difficulty level: Low rank: 1.13 – High rank: 0.63. P-values are derived from a Mann-Whitney test. The changes are counted from one period to another and relate to an increase or a decrease of the average difficulty level of at least one level (e.g. a change from difficulty level 2.0 to 3.1 between period 2 and 3 is counted as one change).

[213] Mean values on mistake share: Low rank - 18.75% - High rank: 27.56%. P-values are derived from a Mann-Whitney test.

[214] Low ranked participants have also required more time to solve problems of difficulty level six but the difference is not significant since there are only 6 observations as compared to 22 observations for high ranked participants. P-values are derived from a Mann-Whitney test.

[215] Cf. Steele (1988); Aronson et al. (1999).

[216] The tests were conducted with different definitions of 'high rank': Rank 1; Rank 1 or 2; Rank 1, 2 or 3.

[217] The mean difficulty level of low ranked participants under unidimen-

Apart from that, multidimensional RPI may have an increased negative impact on the performance of low ranked participants as it can strengthen adverse effects on self-evaluation. More concretely, it potentially involves negative social comparison experiences in different competence areas. This can be illustrated by the finding that nearly all of the overall low ranked participants were also confronted with a low rank regarding their dimension-specific performance.[218] Therefore, it could be argued that multidimensional RPI as compared to unidimensional RPI potentially reinforces negative effects on task motivation and self-efficacy beliefs. Particularly, low ranked participants may have exerted less effort because of decreased motivation since there are no significant differences between the two forms of RPI with regard to strategy changes, mistake share, rank nervousness and interference as well as time required to solve problems of certain difficulty levels.

5.2. Supplemental analyses

5.2.1. Development of performance and attention towards task dimensions over time

In the previous chapter, tests of hypotheses and research questions have used data pooled across the six rounds of the experiment. In order to extend the provided insights on the effects of the experimental manipulations, this section considers how the primary dependent variables performance and attention towards task dimensions have developed over time.

Figure 3 shows the time path of the mean performance of participants by round in all treatments. The graph illustrates that performance differences based on the presence and the dimensionality of RPI already exist in the first round before the participants in the related treatments actually receive RPI. This finding is consistent with previous experimental studies which show that individuals anticipate the potential effect of RPI on their self-evaluation and therefore respond accordingly.[219] In particular, when the performance means are pooled across compensation contracts, they correspond to the results on overall mean performance. Participants' number of points is significantly higher when they are provided with unidimensional RPI (p=0.08) while the positive effect of multidimensional RPI is not significant (p=0.25).[220] However, initially there seem to be no differences between the applied compensation schemes except for a slightly higher level of performance when multidimensional RPI is combined with a flat wage instead of performance-based contract. The moderating influence of the compensation contract under multidimensional RPI appears to strengthen in the following rounds which implies that mean performance is constantly lower in the presence of a performance-based as compared to a flat wage scheme.[221] This insight underpins the results of the previous analyses regarding a negative interaction effect between multidimensional RPI and performance-based compensation.

In contrast to the described differences between the compensation contracts under multidimensional RPI, Figure 3 suggests that performance is after the first round consistently higher in the presence of a performance-based scheme when participants receive unidimensional RPI or no RPI. Further analyses indicate that median performance under unidimensional RPI is higher for flat wage compensation in the first two rounds of the experiment which may explain the negative interaction effect implied by overall median performance and the regression results.[222] The evolution of performance therefore indicates that negative influences of performance-based compensation are relevant in the early rounds while its predicted motivational effects appear to manifest itself in higher performance towards the end of the experiment. This finding could be explained by an increase in practice on the basic multiplication tables that are needed to solve problems of all difficulty levels since this potentially reduces performance declines because of 'choking under pressure'. Specifically, Beilock et al. (2004) argues that 'choking under pressure' may be eliminated in mathematical problem solving as individuals gain practice and can therefore retrieve partial solutions from memory.[223] Importantly, yet, such a countervailing effect of practice on potential 'choking under pressure' in the presence of performance-based compensation cannot be observed with regard to multidimensional RPI.

Overall, the development of performance under unidimensional RPI and performance-based compensation is characterized by some volatility over the first three rounds and a steady increase with regard to the following rounds. Under the flat wage scheme, the development appears to more stable with a slight rise of performance over time. By contrast, the graphs show considerable volatility in performance over all six rounds when participants do not receive RPI or are provided with multidimensional RPI. Remarkably, for both manipulations, the evolution under the two compensation contracts follows the same pattern even though on different total performance levels. Furthermore, there is no considerable directional difference in the evolution of performance under the treatments which do not involve and those that apply multidimensional RPI. The only exception applies to the last round in which the performance under multidimensional

sional RPI equals 2.53. In contrast, the two participants reached an average difficulty level of 1.86 (performance: 87 points) and 1.5 (performance: 23 points, IQR: -1.80).

[218] 16 out of 20 overall low ranked participants under multidimensional RPI were also low ranked on either the 'quantity' or the 'value' dimension of the task (rank 4 or 5). Two low ranked participants even received a low rank on both task dimensions.

[219] Cf. Tafkov (2013); Hannan et al. (2013).

[220] P-values are derived from a Mann-Whitney test. The same result can be obtained from a Pearson Chi-Squared test with regard to median performance (p=0.07, p=0.23). No significant difference in performance can be found between unidimensional and multidimensional RPI in the first round.

[221] Based on the Mann-Whitney test, there is no significant difference in performance under the two compensation contracts in all rounds. Yet, with regard to median performance, the Pearson Chi-Squared test indicates a significant positive effect under a flat wage in round three and four (p=0.05, p=0.09).

[222] There is no significant difference in performance between the two compensation contracts in all rounds.

[223] Cf. Beilock et al. (2004).

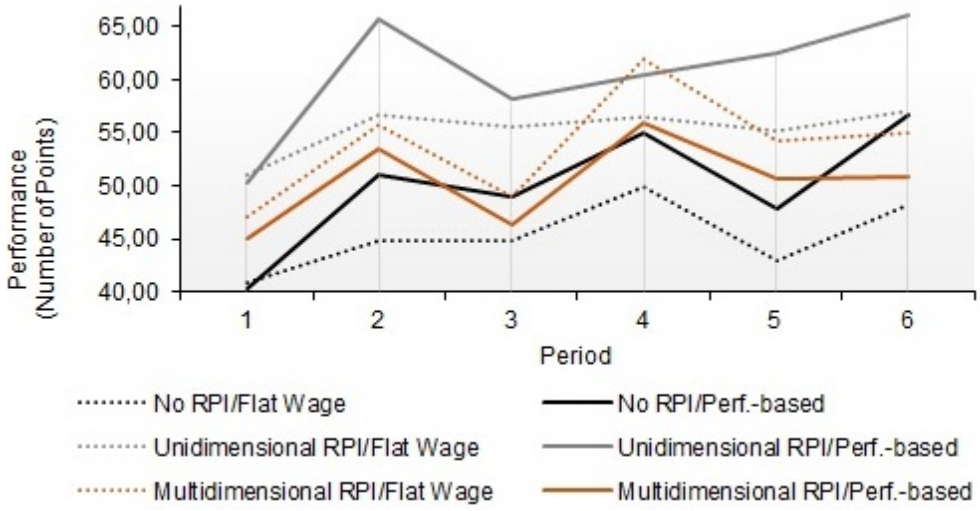

Figure 3: Development of Performance over Periods (Means by Treatment)

slightly decreases while there is a substantially rise for the performance of participants that are not provided with RPI. This result also holds for median performance and can indicate a loss of motivation for participants under multidimensional RPI. Since the provided rankings did not compare current period performance but cumulative performance, they may have not seen the chance to improve their position in the ranking in the last period. Lastly, the development of performance in all treatments reflects an overall increase from the first to the last period which can be related to training and learning effects.[224]

Figure 4 displays the evolution of the mean values per treatment for participants' average difficulty level over all six rounds of the experiment. Broadly, the graphs indicate for nearly all treatments that the average difficulty level constantly fluctuate around a certain baseline with overall changes of no more than 0.58 difficulty level. The only exception is the treatment that does not involve RPI and applies a flat wage since it shows a decrease of 0.84 difficulty levels from the first to the fourth round. While the average difficulty level is initially similar between the two compensation contracts when participants do not receive RPI, this reduction causes an increasing difference since the average difficulty level under performance-based compensation shows a relatively low fluctuation.[225] The development partly complies with the results on performance which show similarity in the first round and a consistently higher level of performance under performance-based compensation in the following rounds. Specifically, it appears that participants provided with a flat wage cannot realize similar performance improvements with the lower difficulty level as those that receive performance-based contract and stay with the higher difficulty level. A possible reason for the decrease of the average difficulty level under flat wage compensation may be the lower level of motivation since solving problems of lower difficulty level requires less cognitive effort.

With regard to the evolution of attention towards task dimensions under unidimensional RPI, Figure 4 first of all shows that the average difficulty level of participants under performance-based compensation is over the whole time the highest among all treatments. Thus, it is from the beginning constantly higher than the average difficulty level under unidimensional RPI and flat wage compensation even though first round performance is similar between both compensation contracts.[226] Notably, the average difficulty level under both compensation schemes fluctuates in the same pattern and seem to be unrelated to the presented stable performance development in the latter rounds of the experiment. Unlike the development under unidimensional RPI, the average difficulty level of participants provided with multidimensional RPI evolves around a comparable baseline under the two compensation contracts. However, while the mean of changes in the average difficulty level over time is lowest for participants provided with a flat wage as compared to all other treatments, there is a considerable volatility under a performance-based scheme.[227] The differences in the development of the average difficulty level are not reflected in the previously shown similar pattern of the evolution of performance under the two compensation contracts. Nevertheless, they might be related to the overall higher performance level under a flat wage scheme. Finally, a comparison between the evolution of the average difficulty level over time

[224]The greatest increases occurred under No RPI/Perf.-based and Unidimensional RPI/Perf.-based. The differences in the mean performance values between the first and last period equal 16.36 and 15.96. For the other treatments, the differences are all below 7.96.

[225]The average difficulty level is significantly higher in round four (p=0.05), five (p=0.05) and six (0.01). The Pearson Chi-Squared test indicates a median difference in round six (p=0.05).

[226]The average difficulty level is significantly higher in round one (p=0.04) and three (p=0.04). The Pearson Chi-Squared test indicates a median difference in round three (p=0.05).

[227]Mean number of changes in the average difficulty level: 1.0. In the remaining treatments, the mean is below 0.88.

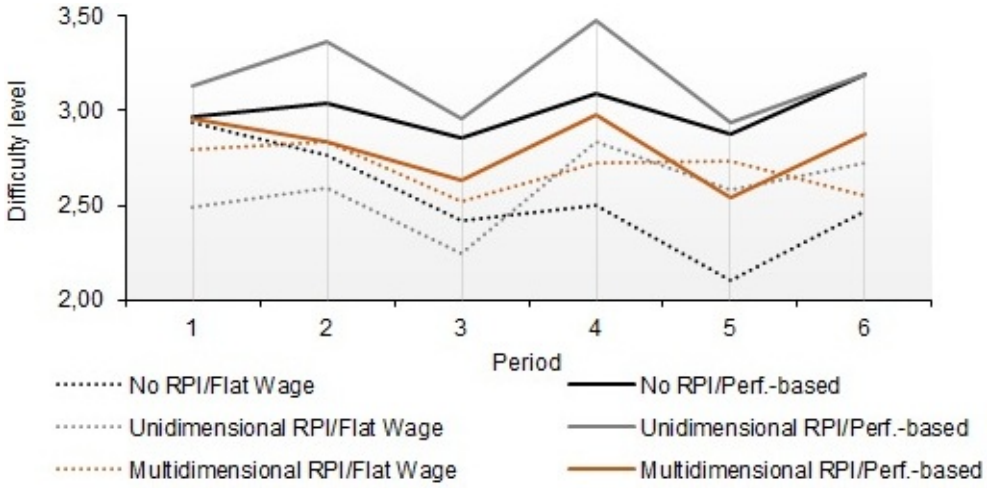

Figure 4: Development of Difficulty Level over Periods (Means by Treatment)

under multidimensional RPI as compared to unidimensional RPI provides no substantiation on differences in learning effects. In this regard, especially the behavioural responses to the provision of RPI in the first round may be relevant which though do not indicate improved adjustments of the average difficulty level under multidimensional RPI.

5.2.2. Post-Experimental questionnaire data

In addition to the observed primary dependent variables of the experiment, post-experimental questionnaire data are used to provide some further insights into the effects of unidimensional and multidimensional RPI.[228] Specifically, it is examined how participants' ranking position relates to their indicated feelings of pride and shame and whether there is a connection between the dimensionality of RPI and statements regarding the development and the use of task-specific strategies.

In the course of the discussion of the results in section 5.1.3, it was argued that social comparison information can adversely affect self-evaluation of low ranked participants and foster negative feelings of inferiority based on a contrastive evaluation of the comparison target. In this context, the data of the post-experimental questionnaire allow to investigate whether participants' position in the rankings created feelings of pride and shame. Using a seven-point Likert scale, participants had to assess how pride they felt about their own performance (1=not at all, 7=to a great degree).[229] With regard to the theoretical predictions, it is expected that the continuous variable for participants' total rank is negatively related to the variable that reflects pride since a smaller rank number constitutes better relative performance. Additional analyses support this supposition since they show that total rank and pride are significantly negatively correlated for both participants that were provided with unidimensional and multidimensional RPI (r= -0.69, p=0.00; r= -0.75, p=0.00).[230] It appears that there is a stronger negative correlation for multidimensional RPI which could be an indication for a strengthening of adverse effects on self-evaluation when participants potentially experience negative social comparison experiences in different competence areas. However, a test on the equality of the correlation coefficients reveals that there is no significant difference. Importantly, no correlation is found for participants that did not receive any form of RPI and were therefore not aware about differences between their own performances and that of their peers.

Apart from that, the post-experimental questionnaire was also used to ask participants about the task strategy that they adopted during the experiment as well as the usefulness of the provided ranking information for finding an optimal strategy. In this regard, it has been hypothesized in the analysis on the behavioural effects of RPI in section 2.1 that multidimensional RPI is more beneficial for learning processes on task strategies than unidimensional RPI. However, so far, the analyses on the observed primary dependent variables did not indicate differences in learning effects between the two forms of RPI. Notably, the post-experimental questionnaire data suggest that multidimensional RPI in general increased the application of a task-specific strategy as compared to unidimensional RPI. More concretely, the number of participants that constituted in the questionnaire that they have not followed any strategy during task execution is significantly higher under unidimensional RPI (p=0.09).[231] Nevertheless, the questionnaire data show that participants which received

[228] See Online-Appendix 5 for the questions of the post-experimental questionnaire.

[229] The question is adapted from Tafkov (2013).

[230] Regarding the additional rankings provided under multidimensional RPI, a significant correlation can be found between pride and the final position in the 'value' ranking (r=-0.47, p=0.00). How- ever, it appears that this is driven by a significant positive correlation between overall rank and 'value' rank (r=0.48, p=0.00).

[231] Participants where asked which strategy they have adopted during task execution and had to choose between the options 'Value Strategy', 'Quantity

multidimensional RPI did not perceive the ranking information to be more useful for strategy development than those that were provided with unidimensional RPI.[232] In addition, there appear to be no differences in the evaluation between the two compensation schemes that were applied with multidimensional RPI. Overall, the participants' average assessment of the usefulness equaled 3.45 (s.d.=2.01) on a seven point Likert scale (1=not at all; 7=to a great degree).

6. Conclusion

The provision of RPI is an important ingredient in the 'gamification' of tasks and is commonly used by firms to increase the performance of their employees. In case employees have to fulfil tasks that involve multiple dimensions, firms have to decide on the dimensionality of RPI and can basically choose between unidimensional and multidimensional RPI. However, little is known about the potential cost and benefits of the provision of different forms of RPI in multidimensional task settings in which overall performance is not only determined by the motivated level of effort but also by the induced attention towards task dimensions. Against this background, the present study has conceptually analysed behavioural effects of unidimensional and multidimensional RPI under different compensation schemes. On this basis, it has applied a controlled laboratory experiment to empirically test the influence of the dimensionality of RPI on performance and attention towards task dimensions.

In sum, the study demonstrates that solely the provision of unidimensional RPI improves performance in multidimensional task settings while the effects for multidimensional RPI are small and insignificant. Importantly, both unidimensional and multidimensional RPI imply a negative effect on the performance of participants with a low position in the provided ranking on overall performance. Additionally, the application of performance-based compensation negatively moderates the performance impact of both forms of RPI which though seem to be especially critical for multidimensional RPI. In this regard, multidimensional RPI induces a significant performance decrease compared to unidimensional RPI in the presence of a performance-based contract. The findings provide some indication that a distortion of attention toward the 'quantity' dimension of the applied experimental task may have caused the negative performance effect. Notably, the experimental results do not indicate increased learning effects regarding a task-specific strategy under multidimensional as compared to unidimensional RPI. Overall, no significant differences in attention towards task dimensions between the two forms of RPI can be proved.

The findings of the study extend the stream of literature that investigates the motivational influence of RPI in task settings with one relevant performance dimension since it enhances the understanding of the effects of RPI in multidimensional task setting. Furthermore, it involves practical implications for the use of RPI in firms which employ multidimensional tasks and have no certain preferences on employees' attention towards task dimensions but rather aim to maximize overall performance. Based on the study's results, it appears to be not beneficial to rank employees based on their dimension-specific performance. In this regard, features of 'gamification' platforms like the possibility to provide employees with symbolic badges for the mastery of different performance dimensions should be used with care.[233] Also, the study emphasizes that firms should not separate the decisions on the implementation of RPI and the use of financial incentives but have to jointly consider possible effects on the motivation of their workforce. Moreover, it is of special relevance regarding the growing adaption of 'gamification' platforms that firms are aware that the provision of RPI can come at the cost of demotivating low ranked employees.

The results of the present study are subject to a number of limitations. As already outlined in chapter 4.1, the generalizability of findings obtained from laboratory experiments is an essential concern.[234] First, results are drawn from a small and homogenous sample of 150 student participants. Second, anonymity was preserved during the experiment and therefore participants did not know how the fictional names used in the rankings matched to the actual persons in the laboratory. This may have lowered the strength of social comparison involvement compared to a workplace context in which colleagues know each other and where 'gamification' leaderboards commonly show photos of employees next to their rank.[235] Also, a lack of identification and psychological closeness to the comparison targets may have promoted the salience of a personal self in opposition to a social self which can enforce negative feelings and demotivation for low ranked participants.[236] Third, the applied experimental task enables a clear test of the underlying theory but is not reflective for multidimensional tasks in the real world as it does not capture all higher-order capabilities and task strategy considerations that might be required from employees to perform well in their jobs. Lastly, in general, it has to be emphasized that the experimental data include various (extreme) outlying observations which reduce the robustness of the results.

Overall, the findings of the study suggest various avenues for further research. Future investigations could shed light on the indicated moderating influence of the applied compensation scheme on the performance effects of RPI in multidimensional task settings. In this context, it seems to be of interest to explore further determinants of varying performance under the two forms of RPI apart from potential

Strategy', 'No Strategy' or 'Other Strategy' In case of a choice of 'Other Strategy', the applied strategy had to be specified. Percentage of participants that chose 'No Strategy': 1) Unidimensional RPI: 20%; 2) Multidimensional RPI: 8%.

[232] There is also no correlation between the perceived usefulness of the ranking information and the performance of the participants under the two forms of RPI.

[233] Cf. Werbach and Hunter (2012); The Economist (2012b)

[234] Cf. Loewenstein (1999); Schnell et al. (2008).

[235] Cf. Werbach and Hunter (2012); Bunchball.

[236] Cf. Stapel and Koomen (2005); Corcoran et al. (2011).

differences in attention towards task dimensions. Especially, potential 'choking under pressure' in the presence of multidimensional RPI may receive attention in connection with the cognitive effects of trade-offs faced by employees when financial rewards are tied to overall performance while competition is facilitated on dimension-specific performance measures. As it was outlined earlier, 'choking under pressure' is a frequent phenomenon in mathematical problems solving and therefore, it appears to be important to additionally test effects with other types of tasks. Furthermore, future research could investigate in greater depth whether multidimensional RPI may enhance adverse performance effects for low ranked employees as it can cause negative social comparison experiences in different competence areas. Finally, since the participants of the conducted experiment were only informed about performance ranks but not about performance levels of their group members, it would be interesting how such additional information may influence performance as well as attention allocation under unidimensional and multidimensional RPI.

References

Acock, C. A. *A Gentle Introduction to Stata*. College Station, 2006.

Antle, R. and Smith, A. An Empirical Investigation of the Relative Performance Evaluation of Corporate Executives. *Journal of Accounting Research*, 24(1):1–39, 1986.

Ariely, D., Gneezy, U., Loewenstein, G., and Mazar, N. Large stakes and big mistakes. *The Review of Economic Studies*, 76(2):451–469, 2009.

Aronson, J., Cohen, G., and Nail, R. P. Self-affirmation theory: An update and appraisal. In *Progress on a pivotal theory in social psychology*, pages 127–147. Eddie Harmon-Jonesand Judson Mills, Washington D. C., 1999.

Azmat, G. and Iriberri, N. The importance of relative performance feedback information: Evidence from a natural experiment using high school students. *Journal of Public Economics*, 94:435–452, 2010.

Baiman, S. Agency Research in Managerial Accounting: A Survey. *Journal of Accounting Literature*, 1:154–213, 1982.

Baiman, S. Agency research in managerial accounting: A second look. *Organizations and Society*, 15(4):341–371, 1990.

Baker, G. Distortion and risk in optimal incentive contracts. *Journal of human resources*, pages 728–751, 2002.

Bandura, A. *Social learning theory*. New York, 1977a.

Bandura, A. Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2):191–215, 1977b.

Bandura, A. and Jourden, F. J. Self-regulatory mechanisms governing the impact of social comparison on complex decision making. *Journal of Personality and Social Psychology*, 60(6):941–951, 1991.

Bandura, A. and Wood, R. Effect of perceived controllability and performance standards on self-regulation of complex decision making. *Journal of Personality and Social Psychology*, 56(5):805–814, 1989.

Baumeister, R. F. Choking under pressure: self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of personality and social psychology*, 46(3):610–620, 1984.

Beach, S. R. H. and Tesser, A. Self-esteem and the extended self-evaluation maintenance model. In Michael, M., editor, *Efficacy, agency, and self-esteem*, pages 145–170. Kernis, New York, 1995.

Beilock, S. L., Kulp, C. A., Holt, L. E., and Carr, T. H. More on the Fragility of Performance: Choking Under Pressure in Mathematical Problem Solving. *Journal of Experimental Psychology*, 133(4):584–600, 2004.

Berger, L., Klassen, K., Libby, T., and Webb, A. Evidence on the Relation between Incentives and Performance in Repeated Tournaments. *Working paper Brock University and University of Waterloo*, 2011.

Blanes i Vidal, J. and Nossol, M. Tournaments without prizes: Evidence from personnel records. *Management Science*, 57(10):1721–1736, 2011.

Bonner, S. E. and Sprinkle, G. B. The effects of monetary incentives on effort and task performance: theories, evidence, and a framework for research. *Accounting, Organizations and Society*, 27(4):303–345, 2002.

Boudreau, K., Helfat, C., Lakhani, K. R., and Menietti, M. E. *Field evidence on individual behavior and performance in rank-order tournaments*. Harvard Business School Working PaperBoudreau2012, 2012.

Bourdieu, P. *La distinction: Critique sociale du jugement*. Paris, 1979.

Bracha, A. and Fershtman, C. Competitive Incentives: Working Harder or Working Smarter? *Management Science*, 59(4):771–781, 2013.

Breusch, T. S. and Pagan, A. R. A simple test for heteroscedasticity and random coefficient variation. *Journal of the Econometric Society*, pages 1287–1294, 1979.

Brickman, P. and Bulman, R. J. Pleasure and pain in social comparison. In *Social comparison processes: Theoretical and empirical perspectives*, volume 149, page 186. 1977.

Brown, D. J., Ferris, D. L., Heller, D., and Keeping, L. M. Antecedents and consequences of the frequency of upward and downward social comparisons at work. *Organizational Behavior and Human Decision Processes*, 102(1):59–75, 2007.

Brown, G. W. and Mood, A. M. On median tests for linear hypotheses. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*. 1951.

Brown, J. L., Fisher, J. G., Sooy, M., and Sprinkle, G. B. The effect of rankings on honesty in budget reporting. *Accounting, Organizations and Society*, 39(4):237–246, 2014.

Brüggen, A. and Moers, F. The role of financial incentives and social incentives in multi-task settings. *Journal of Management Accounting Research*, 19(1):25–50, 2007.

Bunchball. Gamification Case Studies & Customers. URL http://www.bunchball.com/customers. Downloaded: 27.04.2014.

Buunk, B. P. and Mussweiler, T. New directions in social comparison research. *European Journal of Social Psychology*, 31(5):467–475, 2001.

Camerer, C. F. and Hogarth, R. M. The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of risk and uncertainty*, 19(1-3):7–42, 1999.

Carmona, C., Buunk, A. P., Dijkstra, A., and Peiro, J. M. The relationship between goal orientation, social comparison responses, self-efficacy, and performance. *European Psychologist*, 13(3):188–196, 2008.

Castellan, J. N. and Swaine, M. Long-term feedback and differential feedback effects in nonmetric multiple-cue probability learning. *Behavioral Science*, 22(2):116–128, 1977.

Cervone, D., Peake, P. K., and Philip, K. Anchoring, efficacy, and action: The influence of judgmental heuristics on self-efficacy judgments and behavior. *Journal of Personality and Social Psychology*, 50(3):492, 1986.

Charness, G. and Grosskopf, B. Relative payoffs and happiness: an experimental study. *Journal of Economic Behavior & Organization*, 45(3):301–328, 2001.

Charness, G. and Rabin, M. Understanding Social Preferences with Simple Tests. *The Quarterly Journal of Economics*, 117(3):817–869, 2002.

Charness, G., Masclet, D., and Villeval, M. C. The dark side of competition for status. *Management Science*, 60(1):38–55, 2013.

Chung, K. H. and Vickery, D. W. Relative effectiveness and joint effects of three selected reinforcements in a repetitive task situation. *Organizational Behavior and Human Performance*, 16(1):114–142, 1976.

Collins, R. L. For better or worse: The impact of upward social comparison on self-evaluations. *Psychological Bulletin*, 1:51–69, 1996.

Collins, R. L. Among the Better Ones. In *Handbook of Social Comparison*, pages 159–171. Jerry Sulsand Ladd Wheeler, 2000.

Colombino, T., Castellani, S., Grasso, A., and Willamowski, J. Agentville: Supporting Situational Awareness and Motivation in Call Centres. In *From Research to Practice in the Design of Cooperative Systems: Results and Open Challenges*, pages 291–307. Julie Dugdale, 2012.

Congleton, R. D. Efficient status seeking: externalities, and the evolution of status games. *Journal of Economic Behavior & Organization*, 11(2): 175–790, 1989.

Corcoran, K., Crusius, J., and Mussweiler, T. Social comparison: Motives, standards, and mechanisms. *Theories in social psychology*, pages 119–139, 2011.

Croson, R. and Gneezy, U. Gender differences in preferences. *Journal of Economic literature*, pages 448–474, 2009.

Datar, S., Kulp, S. C., and Lambert, R. A. Balancing performance measures. *Journal of Accounting Research*, 39(1):75–92, 2001.

Deci, E. L. and Ryan, R. M. *Intrinsic Motivation and Self-Determination in Human Behavior*. New York, 1985.

Deterding, S., Dixon, D., Khaled, R., and Nacke, L. From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th International Academic MindTrek Conference*, pages 9–15. Envisioning Future Media Environments, 2011.

Deutsch, M. A theory of cooperation and competition. *Human Relations*, 2: 129–152, 1949.

Eisenhardt, K. M. Agency theory: An assessment and review. *Academy of management review*, 14(1):57–74, 1989.

Eriksson, T., Poulsen, A., and Villeval, M. C. Feedback and incentives: Experimental evidence. *Labour Economics*, 16(6):679–688, 2009.

Falk, A. and Fehr, E. Why labour market experiments? *Labour Economics*, 10(4):399–406, 2003.

Falk, A. and Heckman, J. J. Lab experiments are a major source of knowledge in the social sciences. *Science*, 23:535–538, 2009.

Falk, A. and Ichino, A. Clean Evidence on Peer Effects. *Journal of Labor Economics*, 24(1):39–57, 2006.

Falk, A., Dohmen, T., and Sunde, U. Kontrolliert und repräsentativ: Beispiele zur Komplementarität von Labor- und Felddaten. *Perspektiven der Wirtschaftspolitik*, 10:54–74, 2009.

Fehr, E. and Schmidt, K. M. Fairness and Incentives in a Multi-task Principal-Agent Model. *The Scandinavian Journal of Economics*, 106(3):453–474, 2004.

Feltham, G. A. and Xie, J. Performance measure congruity and diversity in multi-task principal/agent relations. *Accounting Review*, pages 429–453, 1994.

Feltovich, N. What's to Know about Laboratory Experimentation in Economics? *Journal of Economic Surveys*, 25(2):371–379, 2011.
Festinger, L. Informal social communication. *Psychological Review*, 57(5): 271–282, 1950.
Festinger, L. A theory of social comparison processes. *Human relations*, 7 (2):117–140, 1954.
Field, A. and Hole, G. J. *How to design and report experiments*. London, 2002.
Fischbacher, U. z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178, 2007.
Frey, B. S. Awards as compensation. *European Management Review*, 4(1): 6–14, 2007.
Garcia, S. M. and Tor, A. Rankings, standards, and competition: Task vs. scale comparisons. *Organizational Behavior and Human Decision Processes*, 102(1):95–108, 2007.
Garcia, S. M. and Tor, A. The n-effect more competitors, less competition. *Psychological Science*, 20(7):871–877, 2009.
Garcia, S. M., Tor, A., and Gonzalez, R. Ranks and rivals: A theory of competition. *Personality and Social Psychology Bulletin*, 32(7):970–982, 2006.
Ginsburgh, V. Awards, success and aesthetic quality in the arts. *Journal of Economic Perspectives*, 17(2):99–111, 2003.
Gist, M. E. and Mitchell, T. R. Self-efficacy: A theoretical analysis of its determinants and malleability. *Academy of Management review*, 17(2): 183–211, 1992.
Goethals, G. R. and Darley, J. M. Social comparison theory: An attributional approach. In *Theoretical and empirical perspectives*, pages 259–278. Social comparison processes, 1977.
Greiner, B. An Online Recruitment System for Economic Experiments. In Macho, K. K. V., editor, *Forschung und wissenschaftliches Rechnen*, pages 79–93. Ges. fuer Wiss. Datenverarbeitung, Goettingen, 2004.
Griffin-Pierson, S. The Competitiveness Questionnaire: A Measure of Two Components of Competitiveness. *Measurement and Evaluation in Counseling and Development*, 23(3):108–115, 1990.
Guimond, S. *Social Comparison and Social Psychology: Understanding Cognition, Intergroup Relations, and Culture*. Cambridge, 2006.
Hannan, L. R., Mcphee, G. P., Newman, A. H., and Tafkov, I. D. *The Effect of Relative Performance Information Detail and Time Horizon on Effort Allocation in a Multi-Task Environment*. Working paper, 2014.
Hannan, R. L., Krishnan, R., and Newman, A. H. The Effects of Disseminating Relative Performance Feedback in Tournament and Individual Performance Compensation Plans. *Accounting Review*, 83(4):893–913, 2008.
Hannan, R. L., Mcphee, G. P., Newman, A. H., and Tafkov, I. D. The Effect of Relative Performance Information on Performance and Effort Allocation in a Multi-Task Environment. *Accounting Review*, 88(2):553–575, 2013.
Hattie, J. and Timperley, H. The power of feedback. *Review of educational research*, 77(1):81–112, 2007.
Hibbard, D. R. and Buhrmester, D. Competitiveness, gender, and adjustment among adolescents. *Sex Roles*, 63(5-6):412–424, 2010.
Hoaglin, D. C., Iglewicz, B., and Tukey, J. W. Performance of some resistant rules for outlier labeling. *Journal of the American Statistical Association*, 81(396):991–999, 1986.
Hogarth, R. M. *Judgement and choice*. The psychology of decision, 1987.
Holmstrom, B. Moral Hazard in Teams. *The Bell Journal of Economics*, 13 (2):324–340, 1982.
Holmstrom, B. and Milgrom, P. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics & Organization*, 7:24, 1991.
Holt, C. A. and Laury, S. K. Risk aversion and incentive effects. *American economic review*, 92(5):1644–1655, 2002.
Johnson, D. W. and Johnson, R. T. *Cooperation and competition*. Theory and research, 1989.
Kachelmeier, S. J., Reichert, B. E., and Williamson, M. G. Measuring and motivating quantity, creativity, or both. *Journal of Accounting Research*, 46(2):341–373, 2008.
Kandel, E. and Lazear, E. P. Peer pressure and partnerships. *Journal of political Economy*, 100(4):801, 1992.
Kerlinger, F. N. and Lee, H. B. *Foundations of Behavioral Research*. 2000.
Kilduff, G. J., Elfenbein, H. A., and Staw, B. M. The psychology of rivalry: A relationally dependent analysis of competition. *Academy of Management Journal*, 53(5):943–969, 2010.
Kluger, A. N. and Denisi, A. The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2):254–284, 1996.
Koenker, R. *Quantile Regression*. Cambridge, 2005.
Koenker, R. and Bassett, G. Regression quantiles, in: Econometrica. *Journal of the Econometric Society*, pages 33–50, 1978.
Ku, G., Malhotra, D., and Murnighan, J. K. Towards a competitive arousal model of decision-making: A study of auction fever in live and Internet auctions. *Organizational Behavior and Human Decision Processes*, 96(2): 89–103, 2005.
Lazear, E. P. and Rosen, S. Rank-Order Tournaments as Optimum Labor Contracts. *Journal of Political Economy*, 89(5):841–864, 1981.
Leary, M. R. and Tangney, J. P. *Handbook of Self and Identity*. 2005.
Levene, H. Robust tests for equality of variances1. In *Contributions to probability and statistics*, volume 2, pages 278–292. Essays in honor of Harold Hotelling, 1960.
Libby, R., Bloomfield, R., and Mark, M. W. Experimental research in financial accounting. *Accounting, Organizations and Society*, 27(8):775–810, 2002.
Locke, E. A. and Latham, G. P. *A theory of goal setting & task performance*. New Jearsy, 1990.
Lockwood, P., Jordan, C. H., and Kunda, Z. Motivation by positive or negative role models: regulatory focus determines who will best inspire us. *Journal of personality and social psychology*, 83(4):854, 2002.
Loewenstein, G. Experimental economics from the vantage-point of behavioural economics. *The Economic Journal*, 109(453):25–34, 1999.
Lynn, M. and Snyder, C. R. *Uniqueness seeking, in: Handbook of positive psychology*. Charles R. Snyder and Shane J. Lopez, Oxford, 2002. 395-410.
Malhotra, D. The desire to win: The effects of competitive arousal on motivation and behavior. *Organizational Behavior and Human Decision Processes*, 111(2):139–146, 2010.
Mann, H. B. and Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, pages 50–60, 1947.
Marquaridt, D. W. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics*, 12(3):591–612, 1970.
May, M. A. and Doob, L. W. Competition and cooperation. *Social Science Research Council Bulletin*, 1937.
Mead, M. *Cooperation and competition among primitive peoples*. New York, 1937.
Meder, M., Plumbaum, T., and Hopfgartner, F. Perceived and actual role of gamification principles. In *Proceedings of the 2013 IEEE/ACM 6th International Conference on Utility and Cloud Computing*, pages 488–493, 2013.
Moldovanu, B. and amd X. Shi, A. S. Contests for Status. *Journal of Political Economy*, 115(2):338–363, 2007.
Morey, N. and Gerber, G. L. Two Types of Competitiveness: Their Impact on the Perceived Interpersonal Attractiveness of Women and Men. *Journal of Applied Social Psychology*, 25(3):210–222, 1995.
Murthy, U. S. The Effect of Relative Performance Information under Different Incentive Schemes on Performance in a Production Task. Working Paper, 2011.
Mussweiler, T. 'Seek and ye shall find': ntecedents of assimilation and contrast in social comparison. *European Journal of Social Psychology*, 31(5): 499–509, 2001.
Mussweiler, T. Comparison processes in social judgment: Mechanisms and consequences. *Psychological Review*, 110(3):472–489, 2003.
Nalebuff, B. J. and Stiglitz, J. E. Prizes and incentives: towards a general theory of compensation and competition. *The Bell Journal of Economics*, pages 21–43, 1983.
Niederle, M. and Vesterlund, L. Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics*, 122:1067–1101, 2007.
Nordstrom, R., Lorenzi, P., and Hall, R. V. A review of public posting of performance feedback in work settings. *Journal of Organizational Behavior Management*, 11(2):101–124, 1991.
Payne, J. W., Bettman, J. R., and Johnson, E. J. *The adaptive decision maker*. Cambridge, 1993.
Poortvliet, P. M., Janssen, O., Yperen, N. W. V., and de Vliert, E. V. Low ranks make the difference: How achievement goals and ranking information affect cooperation intentions. *Journal of Experimental Social Psychology*, 45(5):1144–1147, 2009.
Prendergast, C. The provision of incentives in firms. *Journal of economic*

literature, pages 7–63, 1999.

Reeves, B. and Read, J. L. *Total Engagement: How Games and Virtual Worlds Are Changing the Way People Work and Businesses Compete*. Boston, 2013.

Ryckman, R. M., Hammer, M., Kaczor, L. M., and Gold, J. A. Construction of a Hypercompetitive Attitude Scale. *Journal of Personality Assessment*, 55 (3-4):630–639, 1990.

Ryckman, R. M., Hammer, M., Kczor, L. M., and Gold, J. A. Construction of a personal development competitive attitude scale. *Journal Of Personality Assessment*, 66(2):374–385, 1996.

Rydval, O. and Ortmann, A. How financial incentives and cognitive abilities affect task performance in laboratory settings: An illustration. *Economics Letters*, 85(3):315–320, 2004.

Schmidt, K. M. Social Preferences and Competition. *Journal of Money, Credit and Banking*, 43:207–231, 2011.

Schnell, R., Hill, P. B., and Esser, E. *Methoden der empirischen Sozialforschung*. 2008.

Schreck, P. *Honesty in managerial reporting: How competition affects the benefits and costs of lying, in: Critical Perspectives on Accounting*. 2014.

Sherman, D. K. and Cohen, G. L. The psychology of self-defense: Self-affirmation theory, in: Advances in experimental social psychology. *Advances in experimental social psychology*, pages 183–242, 2006.

Silverman, R. E. Latest Game Theory: Mixing Work and Play, The Wall Street Journal. The Wall Street Journal, 2011. URL http://online.wsj.com/news/articles/SB10001424052970204294504576615371783795248. Downloaded: 27.04.2014.

Simon, H. A. Behavioural economics. *A dictionary of economics*, pages 221–224, 1987.

Simsek, Ö. F. and Yalicetin, B. I feel unique, therefore I am: The development and preliminary validation of the personal sense of uniqueness (PSU) scale. *Personality and Individual Differences*, 49(6):576–581, 2010.

Smith, V. L. Economics in the Laboratory. *The Journal of Economic Perspectives*, pages 113–131, 1994.

Snyder, C. R. and Fromkin, H. L. *Uniqueness: The human pursuit of difference*. New York, 1980.

Sprinkle, G. B. Perspectives on experimental research in managerial accounting. *Accounting, Organizations and Society*, 28(2):287–318, 2003.

Stapel, D. A. and Koomen, W. Competition, cooperation, and the effects of others on me. *Journal of personality and social psychology*, 88(6):1029, 2005.

Stapel, D. A. and Schwinghammer, S. A. Defensive social comparisons and the constraints of reality. *Social Cognition*, 22(1):147–167, 2004.

Steele, C. M. The psychology of self-affirmation: Sustaining the integrity of the self. *Advances in experimental social psychology*, 21(2):261–302, 1988.

Suls, J. E. and Wheeler, L. E. *Handbook of social comparison*. Theory and research, New York, 2000.

Suls, J. E., Martin, R., and Wheeler, L. Social comparison: Why, with whom, and with what effect? *Current Directions in Psychological Science*, 11(5): 159–163, 2002.

Tafkov, I. D. Private and Public Relative Performance Information under Different Compensation Contracts. *The Accounting Review*, 88(1):327–350, 2013.

Tassi, F. and Schneider, B. H. Task-Oriented Versus Other-Referenced Competition: Differential Implications for Children's Peer Relations. *Journal of Applied Social Psychology*, 27(17):1557–11580, 1997.

Taylor, S. E., Wood, J. V., and Lichtmann, R. R. It could be worse: Selective evaluation as a response to victimization. *Journal of Social Issues*, 39(2): 19–40, 1983.

Tesser, A. Toward a self-evaluation maintenance model of social behavior. *Advances in experimental social psychology*, 21:181–227, 1988.

Tesser, A., amd J. C. Collins, N. C., Cornelll, D., and Beach, S. Confluence of self-esteem regulation mechanisms: On integrating the self-zoo. *Personality and Social Psychology Bulletin*, 26(12):1476–1489, 2000.

The Economist. "Gamification", 2012a. URL http://www.economist.com/news/21566581-gamification.

The Economist. More than just a game, 2012b. URL http://www.economist.com/news/business/21565926-video-games-are-behind-latest-fad-management-more-just-game.

TopCoder, 2014. URL http://www.topcoder.com/. Downloaded: 01.05.2014.

Tor, A., Garcia, S. M., and Schiff, T. M. *The Psychology of Competition: A Social Comparison Perspective*. 2013.

Tran, A. and Zeckhauser, R. Rank as an inherent incentive: Evidence from a field experiment. *Journal of Public Economics*, 96(9-10):645–650, 2012.

Triplett, N. The Dynamogenic Factors in Pacemaking and Competition. *The American Journal of Psychology*, 9(4):507–533, 1898.

v. Dijk, F., Sonnemans, J., and v. Winden, F. Incentive systems in a real effort experiment. *European Economic Review*, 45(2):187–214, 2001.

Vorberg, D. and Blankenberger, S. Die Auswahl statistischer Tests und Maße. *Psychologische Rundschau*, 50(3):157–164, 1999.

Wald, A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54(3):426–482, 1943.

Werbach, K. and Hunter, D. *For the Win: How Game Thinking Can Revolutionize Your Business*. Philadelphia, 2012.

Wheeler, L. Motivation as a determinant of upward comparison. *Journal of Experimental Social Psychology*, 1:27–31, 1966.

White, H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Journal of the Econometric Society*, pages 817–838, 1980.

Wills, T. A. Downward comparison principles in social psychology. *Psychological bulletin*, 90(2):245, 1981.

Wood, J. V. Theory and research concerning social comparisons of personal attributes. *Psychological bulletin*, 106(2):231, 1989.

Wood, R. E. and Locke, E. A. Goal setting and strategy effects on complex tasks. *Research in Organizational Behavior*, 12:73, 1990.

Junior Management Science 1 (2016) 1-300

Junior Management Science

journal homepage: www.jums.academy

Practical Implications of the Ambidexterity Concepts

Nina Gusenleitner
Johannes Kepler Universität Linz

Abstract

Scientific publications treating the topic of ambidexterity have experienced a great increase in number since the last twenty years. However, the implications for managers to achieve ambidexterity in practice remain a largely neglected field of research. This thesis aims at bridging the rigor-relevance gap regarding the concept of ambidexterity by systematically reviewing findings from academic and practitioner literature in order to provide practical implications for managers to reconcile exploration and exploitation and to, thus, achieve ambidexterity in practice.

Keywords: Ambidexterity, Practical implications, Exploitation, Rigor-relevance gap

1. Introduction

1.1. Defining the problem and its relevance

In recent years, the discussion about the compatibility of scientific rigor and practical relevance has seemed to gain increasing interest. According to Shapiro et al. (2007), the concept of "getting lost before translation" highlights this problem when referring to the difficulty of connecting scientific findings to practice. This means that the results from scientific research cannot be 'translated' into concrete practical implications for managers due to the fact that researchers and practitioners operate in separate social systems which do not allow cooperation and collaborative research between these two parties (Kieser and Leiner, 2009).

The concept of ambidexterity will be discussed in view of the tendency described above. Ambidexterity is used to explain the capability to change in an organization and includes the balance and linkage of exploration and exploitation. Exploitation can be described as the use, replication and refinement of existing knowledge, whereas exploration is concerned with innovation and the development of new knowledge (Turner et al., 2013). In other words, ambidexterity is the capability to use existing competencies and to develop new competencies simultaneously. The extent to which publications from academic journals and practitioner literature offer practical implications for managers remains unclear.

The topic of ambidexterity is of increasing interest since the last two decades. According to Raisch and Birkinshaw (2008) Raisch and Birkinshaw, "[o]rganizational ambidexterity is currently taking shape as a research paradigm in organizational theory" (Raisch and Birkinshaw, 2008, p. 396). Although there is empirical and theoretical evidence that organizational success can be achieved by means of ambidexterity, there is a lack of understanding concerning the resources which are needed to be ambidextrous and how this can be implemented in practice (Turner et al., 2013). Similarly, O'Reilly and Tushman (2011) argue that "what is missing is a clear articulation of those specific managerial actions that facilitate the simultaneous pursuit of exploration and exploitation" (O'Reilly and Tushman, 2011, p. 8). Birkinshaw and Gupta (2013) address this issue by stating that "if we are to really make a progress on how ambidexterity is achieved, we need much more insight into the nature of managerial capability" (Birkinshaw and Gupta, 2013, p. 293). More precisely, there is a need to clarify the way in which organizations make their decisions, who gets involved in these decisions, how these decisions are implemented, and what these decisions consist of (Birkinshaw and Gupta, 2013).

Research on ambidexterity has focused primarily on the performance effects of structural (e.g., He and Wong, 2004; Lubatkin et al., 2006) or contextual (e.g., Gibson and Birkinshaw, 2004; Adler et al., 1999) ambidextrous organization designs. Additionally, the role of leaders with regard to ambidexterity has been examined by a variety of scholars (e.g., Smith and Tushman, 2005; Jansen et al., 2008). However, the mechanisms for achieving ambidexterity have been largely neglected. This predominant focus on the ambidexterity – performance linkage can be attributed to March

DOI: http://dx.doi.org/10.5282/jums/v1i1pp138-187

(1991) who was the first to distinguish between exploration and exploitation and stated that a combination of these two is essential to ensure the survival of an organization. Birkinshaw and Gupta (2013) bring up an interesting argument concerning the correlation of ambidexterity and performance. They state that if firms engage in both, explorative and exploitative activities, then those firms must, by some degree, outperform those who engage in solely one of those activities. In line with this proposition, Gibson and Gupta (2013) propose that if firms focusing on only one of those activities outperform those focusing on both there must be an incorrect specification of the problem.

In short, what constitutes a major gap in the research of ambidexterity are practical implications for managers which provide mechanisms for achieving ambidexterity in practice and which specify what is needed for the coupling of exploration and exploitation, given that the specialization in only one of these domains is not enough to achieve superior performance. Although prior research on the ambidexterity concept revealed that balancing high levels of exploration and exploitation simultaneously has benefits for a firm's performance (Gibson and Birkinshaw, 2004; He and Wong, 2004), few studies have considered the scientific examination of the drivers of ambidexterity. As it appears to be that the simultaneous pursuit of exploration and exploitation is difficult to achieve in practice, it is surprising that the ambidexterity research lacks the examination of these mechanism that facilitate the achievement of ambidexterity. Building on these findings, the next section will thus try to define a research question with regard to the lack of understanding of how ambidexterity can be achieved in practice.

1.2. Defining a research question

Given that the drivers of ambidexterity remain largely unexplored, the aim of this thesis is to derive practical implications regarding the concept of ambidexterity from scientific publications. The following question should therefore be answered:

> *Which practical implications for managers can be found in academic and practitioner literature to achieve ambidexterity in practice?*

In other words, this thesis aims at providing a summary of mechanisms that should help managers to achieve ambidexterity in an organization. The systematic review of articles from academic journals and practitioner literature should provide a selection of practical implications for managers. The differentiated consideration of the results obtained from either the academic or the practitioner literature should furthermore emphasize differences in how ambidexterity can be achieved in practice. Taylor et al., in 2013(a), have conducted a similar systematic review to investigate the mechanisms for achieving ambidexterity at multiple organizational levels. However, this thesis' aim is not to examine the impact of different mechanisms for achieving ambidexterity on various organizational levels, but on the management which should then be able to take actions accordingly.

Central to the scientific handling of this diploma thesis is the problem of a lack of rigor and relevance in the management field (see e.g., Shapiro et al., 2007; Kieser and Leiner, 2009; Hodgkinson and Rousseau, 2009). Therefore, the main challenge is to investigate whether there are practical implications concerning the ambidexterity concept, or not. If no practical implications can be derived from the prevailing literature and when the theoretical examination of the topic remains dominant, the existence of the rigor-relevance gap concerning the topic of ambidexterity can be approved.

Prior research has primarily focused on examining the antecedents, moderators, and performance outcomes of ambidexterity. Especially, structural, contextual, and leadership-based solutions have seemed to be the major mechanisms for implementing ambidexterity in an organization. This study, however, will take on a different approach in which the management is considered to have an overarching function in implementing different measures for achieving ambidexterity in practice. The systematic review of articles from academic journals and from practitioner literature (i.e., Harvard Business Review, MIT Sloan Management Review, and California Management Review) should help to derive practical implications for managers concerning ambidexterity in order to bridge the rigor-relevance gap.

1.3. Structure of the thesis

This thesis is divided into four main sections, namely a review of the conceptual background concerning ambidexterity, the methodology used for the answering of the research question, the presentation and interpretation of the findings, and the discussion.

First of all, this thesis will analyze the conceptual background on the topic of ambidexterity. With respect to this, a discussion about the rigor-relevance gap in management research will give an introduction of the problematic which is concerned with the connection of scientific findings to practical implications. Specifically, this will help to explain why there exists a rigor-relevance gap concerning the concept of ambidexterity which, in turn, highlights the lack of research regarding the practical implications for achieving ambidexterity in practice. Secondly, the origins and the development of ambidexterity will be described. The analysis of different literature streams should provide an insight into how and under which theoretical lens the topic of ambidexterity has been tackled so far. Furthermore, the different conceptualizations of exploration and exploitation will help to generate a deeper understanding of the matter. Also, the effect of ambidexterity on organizational performance and different environmental factors influencing ambidexterity will be examined. In addition to this, the two main types of ambidexterity and its mechanisms will be explained: structural and contextual ambidexterity. Lastly, an overview of ambidextrous leadership will be given in this section in order to describe the function of leaders and managers in the achievement and development of ambidexterity.

Secondly, this thesis provides an overview of the methodological proceeding for the answering of the research ques-

tion. It will be explained why a systematic literature review is the most appropriate method and what the different steps were in conducting the review. Additionally, the process of inductive category-building and coding, which are both essential for the qualitative analysis of content, will be described. The results of this literature work will be presented in the appendices. Specifically, the appendices include all selected text modules concerning the practical implications of ambidexterity which were obtained from the review of academic and practitioner literature, as well as their corresponding assignment to a category.

Thirdly, the findings obtained from the systematic literature review will be presented in the results section. It will be shown which practical implications concerning the concept of ambidexterity could be derived from academic and practitioner literature. This section is divided into three main categories. Firstly, the findings pertaining to the measures which take place directly at the top management team level will be described and interpreted. These include why managers need to recognize and resolve paradoxical tensions regarding exploration and exploitation, why there is a need for managers to develop an ambidexterity-oriented strategy, why ambidextrous leaders and specific constellations and characteristics of top management teams can facilitate ambidexterity, and which formal and informal mechanisms managers can use to enhance ambidexterity in an organization. Secondly, it will be explained which measures managers can implement within their organization in order to achieve ambidexterity. Particularly, these include structural and contextual arrangements, as well as human resource practices which should facilitate the achievement of ambidexterity in an organization. Lastly, the results section will provide an overview of different moderators and other external factors which influence the attainment of ambidexterity. These are more or less out of scope for decision-making and can often only be marginally influenced by managerial actions. Specifically, it will be shown how ambidexterity can be managed in consideration of the availability of organizational resources, of different environmental factors, and of specific organizational network characteristics. Additionally, it will be explained how ambidexterity can be facilitated through absorptive capacity and dynamic capabilities. The results obtained from either the academic or the practitioner literature will be evaluated separately in order to enable a differentiated appraisal of the results. This should help to assess similarities and differences regarding the practical implications for achieving ambidexterity.

Lastly, in the discussion, the results of this study will be connected to the ambidexterity literature. It will be elaborated whether the research question of this thesis could be answered or not, thus stating if the rigor-relevance concerning the concept of ambidexterity could be bridged or not. The results will be critically discussed and should contribute to the enlargement of the existing knowledge base.

2. Conceptual background

In the last years, the construct of ambidexterity has gained increasing interest and the studies covering this topic have experienced a large increase in number. This led to a multitude of different approaches and a diversity of views held on ambidexterity. It is thus necessary to give an overview of the previous findings on ambidexterity, to comment on the research and to integrate contradictions with regard to the versatility of the concept. Therefore, this section has the aim to review the research done on ambidexterity so far. To do so, conceptual papers as well as leading articles concerning the topic will be reviewed, synthesized, and analyzed. The processing of the current state of research should serve as a base for the specification of the research question. The later answering of the research question should show which practical implications can be derived from the ambidexterity literature.

The section on the conceptual background will give an overview of the roots of the ambidexterity concept and the associated learning modes, exploration and exploitation. To do so, landmark articles, such as the one of Duncan in 1976 and March (1991), will be discussed as they were the first to introduce the notions of ambidexterity, exploration, and exploitation. In addition to this, the examination of leading articles from the mid-nineties to 2004 (e.g., Tushman and O'Reilly, 1996; Gibson and Birkinshaw, 2004; He and Wong, 2004) will help to define the ambidexterity concept and to put its different aspects into a theoretical framing. Review articles (e.g., Raisch and Birkinshaw, 2008; Turner et al., 2013) will be used to connect the findings, to filter out contradictions, and to give additional inputs on certain ideas. But first of all, the discussion about the rigor-relevance gap in management research will explicate the problem of translating the scientific findings of the concept of ambidexterity into practical implications for managers. These findings will be revisited in the end of this section in order to show why the definition of practical implications concerning ambidexterity still constitutes a major gap in this field of research.

2.1. The rigor-relevance gap in management research

The rigor-relevance debate in management research gives rise to the questions if and how the rigor-relevance gap can be bridged. Central to this debate is the struggle of connecting scientific findings to practical situations, of connecting scientific rigor to practical relevance. In order to explain this topic, two opposing articles will be discussed. The first article assumes that the rigor-relevance gap in management research is unbridgeable (Kieser and Leiner, 2009), while the second article suggests that the rigor-relevance gap can be overcome (Hodgkinson and Rousseau, 2009).

Assuming that the rigor-relevance gap is unbridgeable, then Shapiro et al. (2007) notion of getting lost "before translation" applies. This means that the results gained from scientific research are unconnectable to and, thus, untranslatable for practice. According to Kieser and Leiner (2009) this problem does not only arise from different languages and

styles in the scientific community, but also from different logics, meaning different ways in defining and managing problems. They base this assumption on Luhmann's system theory which implies that social systems are autopoietic. Autopoietic means that communication elements of one system cannot be integrated into the communication of other systems because the systems are autonomous and isolated from one another.

Despite these seemingly unbridgeable communication problems, some authors suggest that the cooperation between practitioners and management researchers can help the problem. However, Kieser and Leiner (2009) counter this assumption by stating that social systems can only irritate each other in such a way that the other systems are coerced to respond. From this irritation, however, there might arise inspiration. Inspiration in this sense means that bilingual or bi-competent facilitators should provide descriptions of practical situations for researchers so that they can deduce relevant science concepts and should deliver interpretations that practitioners might consider inspiring (Kieser and Leiner, 2009). Although Kieser and Leiner do not regard collaboration as a solution to bridging the rigor-relevance gap, they note that "[o]nly with training in theory and methodology can practitioners collaboratively produce scientific knowledge with researchers" (Kieser and Leiner, 2009, p. 528). However, they criticize that it is doubtful whether the training of practitioners would lead to better research that is also perceived as useful by practitioners who did not participate in the research collaboration.

In contrast to the article of Kieser and Leiner (2009), Hodgkinson and Rousseau claim that bridging the rigor-relevance gap in management is indeed possible and that "such bridging is increasing; and, as such, increasingly normal, encouraged, and legitimated" (Kieser and Leiner, 2009, p. 535). While Kieser and Leiner (2009) argue that science is overused, too specialized, and not suitable for solving realistic management problems, Hodgkinson and Rousseau (2009) counter these assumptions by claiming that management education often relies on cases with little use of scientific principles, that it is provided by non-researchers, and that it can, if it is appropriately science-based, be a basis for successful management.

Furthermore, Hodgkinson and Rousseau (2009) state that Kieser and Leiner (2009) grounding, the system theory, lacks empirical support. They disagree with the assumption that all systems are autopoietic and suggest that collaborations between researchers and practitioners can be developed with the support of an appropriate training in theory and research methods. In turn, these collaborations can lead to both, high quality scholarship and social usefulness. The knowledge transfer between scholars and practitioners is enabled by dissemination, practice, reflection, and "in the form of knowledge artefacts such as psychometric tests, scenario planning tools, and management science algorithms" (Hodgkinson and Rousseau, 2009, p. 539). Hodgkinson and Rousseau (2009) do also not agree with Kieser and Leiner (2009) concern that the processes of the prime institution would be undermined due to the different objectives that academic and practitioner institutions pursue. They rather suggest that the growing interest in converging scientific rigor and practical relevance leads to a higher level of triangulation of the results which, in turn, promotes a deeper understanding of different phenomena in an organization.

Despite these differences concerning the question of whether the rigor-relevance gap in management is bridgeable or not, both Kieser and Leiner (2009) and Hodgkinson and Rousseau (2009) agree that practitioners and researchers often use fundamentally different categories to describe certain phenomena of their institution. Accordingly, the language is not the only problematic in translating findings from science into practice and vice versa. Moreover, researchers as well as practitioners both "need opportunities to act in ways that bring them in new perspectives and competencies which they in turn use to attain their goals more effectively" (Hodgkinson and Rousseau, 2009, p. 543). Only by changing their perspectives, scholars and practitioners can collaborate with each other and build closer ties between research and practice.

Linking these insights of the rigor-relevance gap to ambidexterity gives rise to a multitude of different leverage points. Ambidexterity is an abstract, impalpable concept. Some firms might act ambidextrous without even knowing it, or even without knowing what ambidexterity is. In other words, practitioners or managers might not be aware of the ambidexterity concept but their organization might still be ambidextrous. The intangibility of ambidexterity might therefore hamper the formulation of concrete implications for practitioners. The following sections will elucidate different views of approaching the concept of ambidexterity and explain its moderators, performance outcomes, and antecedents. The theoretical framing of ambidexterity will serve as a base for narrowing down the topic in order to highlight the necessity of defining implications for achieving ambidexterity in practice which still constitutes a major gap in this field of research.

2.2. The origins and the development of ambidexterity

The roots of ambidexterity can be found in the work of Duncan in the year 1976. Duncan was basically the first to deploy the term organizational ambidexterity and highlights the importance of dual structures. Dual structures can be achieved by "monodextrous" units that spatially separate exploratory from exploitative activities (Duncan, 1976; as cited in Raisch and Birkinshaw, 2008). Dual structures can be related to the concept of structural ambidexterity, meaning the spatial separation of exploration and exploitation. McDonough and Leifer (1983) present a "counter-concept" which involves the implementation of "parallel structures". Parallel structures can be related to the concept of contextual ambidexterity, whereby people can switch between exploitative and explorative activities and do not solely focus on one of the two activities (McDonough and Leifer, 1983).

The work of March (1991) is considered to be another pioneering article in the examination of ambidexterity. Con-

sidering the stream of organizational learning, March used the terms exploration and exploitation to explain that the balance of these two is essential for a system to survive and grow. Exploration can be related to terms, such as "search, variation, risk taking, experimentation, discovery, or innovation", whereas exploitation comprises elements like "refinement, production, efficiency, selection, and implementation" (March, 1991, p.71). March (1991) highlights the importance of establishing a balance between exploration and exploitation for two reasons. Firstly, the extensive use of exploration (failure trap) should not outplay exploitation, because high investments in innovation may not lead to long term gains when the focus lies on the exploration of new alternatives without the true improvement of competencies (pp. 71-72). Secondly, and conversely, the exclusive use of exploitation (success trap) may lead to inertia and the inability to adapt to changing environments when experimentation is neglected due to the success experienced through the refinement of skills in the course of exploitation (pp. 71-72). The long-term success of an organization can thus be assured by the appropriate use of exploitation to guarantee viability and the simultaneous use of exploration to guarantee success in the future (March, 1991).

2.2.1. Examining ambidexterity through different literature streams

One difficulty of connecting different elements and variables of organizational ambidexterity with each other lies in the multitude of research streams under which the concept of organizational ambidexterity is explored (Raisch and Birkinshaw, 2008). Since 1996, when various authors (e.g., Tushman and O'Reilly, 1996; Adler et al., 1999; Gibson and Birkinshaw, 2004; He and Wong, 2004) have contributed to the conceptualization of the ambidexterity concept, research has switched its focus. The proliferation of articles on ambidexterity since 2005 has led to a more specific examination of the topic, including issues such as the consequences, antecedents, or moderators of ambidexterity. Raisch and Birkinshaw (2008) screen the concept of ambidexterity through five different literature streams: organizational learning, technological innovation, organizational adaptation, strategic management, and organizational design. The main findings gathered from examining ambidexterity through various literature streams will be explained below with the aim of providing an overview of how differently the topic of ambidexterity has been tackled so far.

Organizational learning

Concerning the research stream of *organizational learning*, Raisch and Birkinshaw (2008) adduce March (1991) in which he explains in how far and in which way learning can be related to either exploitation or exploration. In general, organizational learning is based on routines (Güttel and Konlechner, 2009). Exploring and exploiting organizational knowledge results in learning which is embedded in organizational routines. Tushman and O'Reilly (1996) state that organizational learning is necessary to achieve congruence between the organizational strategy, structure, culture, and people and the (positive) changes in the market.

Technological innovation

Raisch and Birkinshaw (2008) have found that when considering the research stream of *technological innovation* two types of innovation seem to be predominant. Firstly, the authors refer to incremental innovation which involves small changes and alternations in existing products and concepts. Smith and Tushman (2005) relate incremental innovation to exploitative activities which are used to respond to existing customer demands. Secondly, Raisch and Birkinshaw (2008) identified radical innovation as the other dominant type of technological innovation where change is fundamental and involves the replacement of existing products or business concepts with new ones. Smith and Tushman (2005) relate this type of innovation to explorative activities in order to meet the needs of novel customers. Tushman and O'Reilly (1996) highlight the technology cycle which involves the proliferation of a dominant design, the substitution of this design through competition, and lastly the proliferation of a new standard. From this cycle, the authors derive that organizations and mangers must steadily adapt and realign their strategies in view of the fundamentally changing technologies.

Organizational adaptation

With regard to the literature stream of *organizational adaptation*, Raisch and Birkinshaw (2008) found that a majority of researchers highlight the necessity of a balance between continuity and change. In this context, Tushman and O'Reilly (1996) argue that the pursuit of exploitative activities during evolutionary change and the emphasis on exploration during revolutionary and radical change is essential for an organization to be successful. In other words, organizations and managers have to be able to implement incremental and revolutionary change, meaning they have to be ambidextrous to be successful in the long run. Tushman and O'Reilly (1996) explain that the achievement of ambidexterity requires strong organizational and managerial skills to establish oneself in a mature market (exploitation) as well as to innovate in products (exploration). According to the authors, exploitation includes factors, such as "cost, efficiency, and incremental innovation", whereas exploration comprises things like "radical innovation, speed, and flexibility" (p.11). Tushman and O'Reilly (1996) state that organizations who manage to adapt to changing environmental conditions during periods of change are most likely to be successful. They base this assumption on the Darwinian evolutionary theory and the adjunctive processes of variation, selection, and retention. Raisch and Birkinshaw (2008) also mention other concepts related to organizational adaptation, such as organizational identity, absorptive capacity, and organizational routines. The authors conclude that the main finding of the constructs mentioned above implies that a multitude of (too radical) change activities lead to organizational chaos, whereas to little change actions lead to inertia. Therefore, an organization constantly needs to

adapt its explorative and exploitative activities with regard to changes in the environment.

Strategic management and dynamic capabilities

In order to examine the findings on ambidexterity in the research stream of *strategic management*, Raisch and Birkinshaw (2008) refer to the internal ecology model of strategy making by Burgelman (1991, 2002). According to this model, there are two processes which can be related to either exploration or exploitation. The first process is the induced strategic process which focusses on the use of already existing knowledge and can therefore be related to exploitation. The second process is the autonomous strategic process including the creation of new skills and competencies which can be related to exploration. Similar to many other scholar's findings, Burgelman (1991) argues that organizational success can be achieved through a balance between the two processes of exploration and exploitation.

Dynamic capabilities of a firm are closely linked to the concept of ambidexterity and have its roots in the strategic management domain. Therefore, the topic of dynamic capabilities and its connection with ambidexterity will be briefly addressed in this section. Various scholars (e.g., He and Wong, 2004; O'Reilly and Tushman, 2008) refer to the notion of dynamic capabilities when trying to explain the balance between exploration and exploitation. The linkage of dynamic capabilities and ambidexterity has also been stressed by other researchers which discuss meta-capabilities (Gibson and Birkinshaw, 2004) or meta-routines (Adler et al., 1999) in this regard. Teece, Pisano, and Shuen define dynamic capabilities as "the firm's ability to integrate, build, and reconfigure internal and external competences to address rapidly changing environments" (Teece et al., 1997, p. 516). O'Reilly and Tushman (2008) conceptually describe the connection between dynamic capabilities and ambidexterity basing their assumptions on the findings of Teece (2007), who stresses "orchestration processes" which include learning, reconfiguration, coordination, and integration. According to O'Reilly and Tushman (2008), ambidexterity can be defined as an organization's core dynamic capability because of the concurrent integration of exploration and exploitation. However, this alone does not constitute a competitive advantage yet. It is the configuration of resources which, in turn, may result in a competitive advantage. O'Reilly and Tushman also stress that "the ability of senior managers to seize opportunities through the orchestration and integration of both new and existing assets to overcome inertia and path dependencies is at the core of dynamic capabilities" (O'Reilly and Tushman, 2008, p. 187). To conclude, dynamic capabilities enable a firm to exploit existing competencies and to simultaneously explore new competencies as well as to reconfigure organizational resources in order to seize existing and emerging opportunities (O'Reilly and Tushman, 2008). Ambidexterity is thus a dynamic capability as the simultaneous pursuit and reconfiguration of exploration and exploitation processes allows firms to adapt to changing environments.

Organization design

Regarding *organization design*, Raisch and Birkinshaw (2008) emphasize prior research (e.g., Burns & Stalker, 1961) concerning organizational features which should enable exploration and exploitation. The two most predominant solutions with regard to organization design seem to be the implementation of either mechanistic or organic structures in order to facilitate ambidexterity (Burns & Stalker, 1962; as cited in Raisch and Birkinshaw, 2008). Mechanistic structures involve elements, such as "standardization, centralization, and hierarchy" in order to enable efficiency (exploitation) (p. 380). Organic structures provide "high levels of decentralization and autonomy" to encourage flexibility (exploration) (p. 380). The examination of ambidexterity through diverse literature streams led to a multitude of different conceptualizations of the concept. This constitutes a problem in that there is no common understanding of how ambidexterity should be defined, as well as how exploration and exploitation interact with each other. Of course this problematic makes it even more difficult to understand the already abstract concept of ambidexterity. The section below will provide insights into how ambidexterity, or exploration and exploitation, have been conceptualized so far. This should contribute to the specification of this thesis' problem and research question.

2.3. Conceptualizing ambidexterity

In general, organizational ambidexterity refers to a firm's ability to pursue two different, or even contrasting, things simultaneously. Tushman and O'Reilly define organizational ambidexterity as "the ability to simultaneously pursue both incremental and discontinuous innovation [...] from hosting multiple contradictory structures, processes, and cultures within the same firm" (Tushman and O'Reilly, 1996, p. 24). This assertion is equally in line with March (1991) proposition that a balance between exploration and exploitation is essential to assure a firm's survival. In other words, for an organization to act ambidextrously there needs to be an alignment between its current procedures and its ability to adjust to changing environmental conditions (Gibson and Birkinshaw, 2004).

The coupling of exploration and exploitation is vital for a firm to be ambidextrous. However, exploration and exploitation are still two different, and often contradictory, constructs which need a different approach respectively (O'Reilly and Tushman, 2013). He and Wong put this assumption forward by claiming that "exploration and exploitation require substantially different structures, processes, strategies, capabilities, and cultures to pursue and may have different impacts on firm adaptation and performance" (He and Wong, 2004, p. 481). The authors link exploration to "organic structures, loosely coupled systems, path breaking, improvisation, autonomy and chaos, and emerging markets and technologies", whereas exploitation is linked with "mechanistic structures, tightly coupled systems, path dependence, routinization, control and bureaucracy, and stable markets and technologies" (p. 481).

Since March's pioneering article of 1991, the terms of exploration and exploitation have been reused by many other scholars in the fields of, for example, technological innovation, organizational learning, organizational design, or organizational adaptation (Gupta et al., 2006). However, the definitions and conceptualizations of these two dimensions seem to be inconsistent. Gupta et al. note that "[t]there appears to be a consensus around the view that exploration refers to learning and knowledge (i.e., the pursuit and acquisition of new knowledge). However, a similar consensus is lacking on the question of whether exploitation refers solely to the use of past knowledge or whether it also refers to the pursuit and acquisition of new knowledge, [...]" (Gupta et al., 2006, p. 693). They infer that, based on March (1991), there is at least some learning involved in every activity and that it is therefore logical to differentiate exploration and exploitation by considering the amount or type of learning rather than if learning is existent or not.

This leads to the conclusion that the terms exploration and exploitation are not used consistently by scholars when describing the concept of ambidexterity. Gibson and Birkinshaw (2004), for example, speak of alignment and adaptability when referring to the two activities of exploration and exploitation, Adler et al. (1999) use the terms efficiency and flexibility, and Tiwana (2008) refers to the two dimensions as strong ties and bridging ties. Others again speak of explorative and exploitative knowledge sharing (e.g., Im and Rai, 2008), centrifugal and centripetal forces (e.g., Sheremata, 2000), or incremental and discontinuous innovation (e.g., O'Reilly and Tushman, 2004; Tushman and O'Reilly, 1996). The terms exploration and exploitation, however, have achieved the greatest popularity in scholarly articles and this terminology will therefore be retained in this thesis.

2.3.1. Different dimensions of ambidexterity

Apart from the different designations of exploration and exploitation, their conceptualization also varies. Once, exploration and exploitation are seen as orthogonal, another time they are described as two poles on a continuum (Gupta et al., 2006). The conceptualization of orthogonality or continuity is linked to two different mechanism, either punctuated equilibrium or ambidexterity, with which a balance between exploration and exploitation can be achieved. Punctuated equilibrium is the better solution when exploration and exploitation are conceptualized as two mutually exclusive ends of a continuum. Ambidexterity should be used when exploration and exploitation are viewed orthogonal. The explicit differentiation of these ambidexterity dimensions should offer greater precision to the conceptualization and operationalization of the construct, and provide a basis on which to analyze the practical implications for managers for achieving ambidexterity in practice.

Distinguishing continuity from orthogonality

Gupta et al. (2006) explain that exploration and exploitation are likely to be mutually exclusive, meaning conceptualized as two ends of a continuum, when they compete for scarce resources. However, organizational resources are not always finite and the availability of external resources can help the constraint of scare internal resources (Gupta et al., 2006). Figure 1 depicts exploration and exploitation as two ends of a continuum. The U-shaped curve implies that whenever one of the two activities is pursued at a high level, the other one can only be pursued at lower levels. On the other hand, Figure 1 illustrates exploration and exploitation as orthogonal which means that high levels of either exploration or exploitation in one domain may simultaneously be present with high levels of exploration or exploitation in the other domain. Orthogonal thus means that exploration and exploitation are not competing with each other but that the two can coexist. As an example of the conceptualization of exploration and exploitation as orthogonal, Gupta et al. (2006) mention the case of Cisco where the product R&D is connected with manufacturing, sales, and service via relatively standardized interfaces. This loose coupling of the domains of exploration and exploitation makes it possible to simultaneously achieve both and thus exploration and exploitation can be described orthogonal. The scarcity of resources is, unlike in the conception of exploration and exploration as two poles of a continuum, of no interest when speaking of orthogonality (Gupta et al., 2006).

Furthermore, the level of analysis, meaning whether ambidexterity is analyzed on the organizational, the business unit, or the individual level, determine the conceptualization of the balance between exploration and exploitation (Gupta et al., 2006). Accordingly, Gupta et al. (2006) note that the simultaneous pursuit of exploration and exploitation might be easier for a group or an organization than it is for an individual. On individual levels, it may be more difficult to simultaneously pursue both activities because individuals might not have the appropriate change routines, the possibility of labor division, and a management which controls the allocation of resources and reacts to necessities of change (Gupta et al., 2006). The conclusion from this is that, within a single domain (i.e. an individual or a subsystem), exploration and exploitation will probably be conceptualized as two poles on a continuum, thus being mutually exclusive. Across different, loosely coupled domains (i.e. multiple individuals or subsystems), on the other hand, exploration and exploitation will be orthogonal.

In conclusion, Gupta et al. (2006) argue that the relationship between exploration and exploration depends on whether the two compete for scarce resources and on the level of analysis, meaning a single or multiple domains. For researchers this implies that they need to choose between continuity and orthogonality based on these two factors. When the premise is continuity, then "the correct test for the beneficial effects of balance would be to test for an inverted U-shaped relationship between the degree of exploration (or exploitation) and organizational performance" (p. 697). When orthogonality is the premise, then "the correct test for the beneficial effects of balance would be to test for a positive interaction effect of the two types of learning on organizational performance" (p. 697).

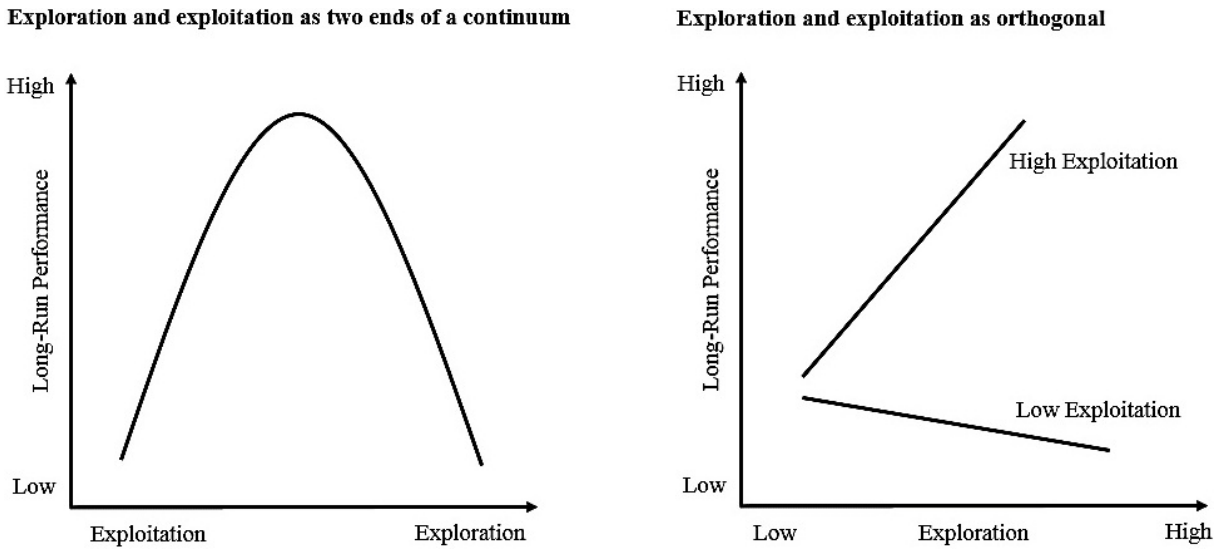

Figure 1: Conceptualizations of exploration and exploitation and its effect on performance (adopted from Gupta et al., 2006, p. 697).

Distinguishing ambidexterity from punctuated equilibrium

General agreement exist on the need for balance between exploration and exploitation. The discussion about how organizations should pursue exploration and exploitation remains open and is treated differently by various scholars (e.g., Junni et al., 2013; Burgelman, 2002; Gupta et al., 2006; Smith and Tushman, 2005; Tushman and O'Reilly, 1996). Gupta et al. (2006) identify two mechanisms that should help to achieve this balance: ambidexterity and punctuated equilibrium. While ambidexterity includes the simultaneous pursuit of exploration and exploitation, these two domains are pursued sequentially within punctuated equilibrium.

As already mentioned earlier, ambidexterity can be defined as the ability of pursuing exploration and exploitation simultaneously. There are two main types of organizational ambidexterity: contextual and structural ambidexterity. Contextual ambidexterity includes the simultaneous pursuit of exploration and exploitation within a subsystem (i.e. an individual or business unit) (see Adler et al., 1999; Gibson and Birkinshaw, 2004). Gibson and Birkinshaw define contextual ambidexterity as "the behavioral capacity to demonstrate alignment and adaptability across an entire business unit" (Gibson and Birkinshaw, 2004, p. 209). Structural ambidexterity includes the simultaneous pursuit of exploration and exploitation across different subsystems, meaning across individuals, business units, or even organizations (Tushman and O'Reilly, 1996). Tushman and O'Reilly describe structural ambidexterity as the "ability to simultaneously pursue both incremental and discontinuous innovation that result from [...] hosting multiple contradictory structures, processes, and cultures, within the same firm" (Tushman and O'Reilly, 1996, p. 24). Although explorative and exploitative units operate separately, they still assure a simultaneous pursuit of both activities within an organization.

Punctuated equilibrium is the second possible mechanism for balancing exploration and exploitation. Contrary to structural or contextual ambidexterity which both involve the simultaneous pursuit of exploration and exploitation, the punctuated equilibrium approach follows a cyclical or sequential pursuit of the two domains within a subsystem. This suggests that periods of exploration (revolutionary change) alternate with periods of exploitation (evolutionary change) (Burgelman, 2002). Thus, punctuated equilibrium involves the "cycling between long periods of exploitation and short bursts of
exploration" (Gupta et al., 2006, p. 698) which implies that the focus is sequentially distributed to the divergent goals of either explorative or exploitative activities. The idea of temporal separation roots in the assumption that "the mindsets and organizational routines needed for exploration are radically different from those needed for exploitation, making a simultaneous pursuit of both all but impossible" (Gupta et al., 2006, p. 695).

Showing that ambidexterity and punctuated equilibrium both are viable, yet different mechanisms for achieving a balance between exploration and exploitation, Gupta et al. (2006) raise the question of whether these two mechanisms can be substituted with one another or whether their implementation depends on the environmental or organizational context. They argue that "[i]f one is analyzing exploration and exploitation within a single domain (i.e., an individual OR a subsystem), and exploration and exploitation are rightly conceptualized as the mutually exclusive ends of a continuum, ambidexterity is simply not an option, and the individual or subsystem must resort to punctuated equilibrium. [...] In contrast, [...] if one is analyzing exploration and exploitation in multiple, loosely connected domains, the two become orthogonal tasks, and it becomes entirely feasible (and perhaps desirable) to pursue ambidexterity" (Gupta et al., 2006, p. 698). In conclusion, this means

that tightly coupled systems would require punctuated equilibrium rather than ambidexterity. However, if individuals or subsystems are loosely coupled with each other, ambidexterity would provide the best possibility for long-term adaptation across the single domains.

A variety of authors (e.g., Simsek et al., 2009; Raisch and Birkinshaw, 2008) consider punctuated equilibrium as one type of ambidexterity. In this thesis, however, ambidexterity will exclusively be conceptualized as the simultaneous pursuit of exploration and exploitation within or across individuals and/or subunits. Approaches concerning "sequential ambidexterity", meaning the temporal switching between explorative and exploitative activities, will therefore be neglected.

2.3.2. Why a balance is necessary for achieving ambidexterity

The activities required for ambidexterity, exploration and exploitation, each compete for the same and/or scarce resources which requires a trade-off between these two (Simsek et al., 2009; Turner et al., 2013; He and Wong, 2004). Birkinshaw and Gupta suggest that "a central part of what firms do is manage the tensions that exist between competing objectives; that is, they seek to achieve some form of ambidexterity" (Birkinshaw and Gupta, 2013, p. 290). More precisely, managerial competence is needed to make trade-offs and to allocate resources according to competing demands (Birkinshaw and Gupta, 2013). So, with regard to ambidexterity, managers need to make trade-offs between the competing domains of exploration and exploitation and have to decide which resources are needed to pursue each of the two activities. The possible synergistic effect between exploration and exploitation requires managers to handle the balance between the two (He and Wong, 2004).

Given the basic conflicts and tensions which occur through the simultaneous use of exploration and exploitation, Gupta et al. (2006) refer to March (1991) assumption that exploration and exploitation are fundamentally incompatible and therefore firms are susceptible to being trapped in either a success or a failure trap. A success trap emerges when exploitation leads to success in the early stages and when this success seduces a firm to exploit even more in this seemingly successful area (Gupta et al., 2006). Contrary to the success trap, a failure trap arises when exploration leads to failure and when firms try to overcome this failure by exploring even more, meaning to reinforce innovation (Gupta et al., 2006). This implies that new ideas are replaced by other innovations without being fully developed and thus do not contribute to the organizational performance. He and Wong (2004) explain the success and the failure trap in more detail. On the one hand, if firms engage in experimentation and innovation (exploration) they risk not being able to improve and refine existing competencies (failure trap). On the other hand, when firms try to adapt to given environmental conditions (exploitation) they risk being inert and not able to respond to changing environments and new opportunities (success trap). In addition to this, failed explorative activities may lead to losses in a firm's successful routines in existing fields which cannot be compensated with the experimentation set into operation.

Tushman and O'Reilly state that long-term success can be achieved through the "alignment of strategy, structure, people, and culture" during periods of evolutionary change (Tushman and O'Reilly, 1996, p. 11). When organizations go through periods of revolutionary change, the alignment of these factors needs to undergo a shift at the same time (Tushman and O'Reilly, 1996). However, as organizations grow, their structures and systems solidify and it gets more difficult to implement change. Tushman and O'Reilly (1996) call this phenomenon the "success syndrome", which was earlier discussed as success trap, and distinguish between two types of inertia which impede change actions: structural and cultural inertia. Structural inertia emerges from existing interdependences and complexities in an organization, whereas cultural inertia results from the age and success of an organization which inhibits the alternation of institutionalized norms and values (Tushman and O'Reilly, 1996). The success syndrome can thus be explained by a situation of success in a stable market which is disrupted by a shift in the market. This leads, through cultural or structural inertia, to the failure to adapt to new conditions. In other words, the extensive focus on exploitation of existing competencies at the expense of exploration of new alternatives will lead to a success trap. The performance of an organization will thus suffer in the long rung when inertia inhibits an organization to adapt to changing environments (Smith and Tushman, 2005).

In addition to the short-term alignment of strategy, structure, and culture managers may need to "cannibalize" their own business by destroying the established alignment in order to adapt to changing competitive markets or technologies (Tushman and O'Reilly, 1996). Due to the quickly changing environmental demands, He and Wong (2004) similarly argue that there needs to be a "counterbalance" between exploration and exploitation. This implies that firms should not try to create an equilibrium, but should rather focus on the disruption of the balance between exploration and exploitation. In sum, this means that exploration and exploitation need to be reconciled so that one activity is not overwhelmed by the other.

Approaches to managing ambidexterity

As just mentioned above, exploration and exploitation are two different and sometimes even contradictory agendas which need to be balanced and managed efficiently to ensure the long-term success of a firm. Birkinshaw and Gupta (2013) provide suggestions on how firms can position themselves in order to invest in appropriate amounts necessary for exploration and exploitation and, thus, to make trade-offs between these two agendas. They illustrate different approaches to managing ambidexterity with the help of a curve (see Figure 2), which is based on Porter's (1996; as cited in Birkinshaw and Gupta, 2013) *efficiency frontier*.

According to Figure 2, firms are likely to place themselves somewhere on the curve, the efficiency frontier, in order to invest in exploration and exploitation. When firms want to

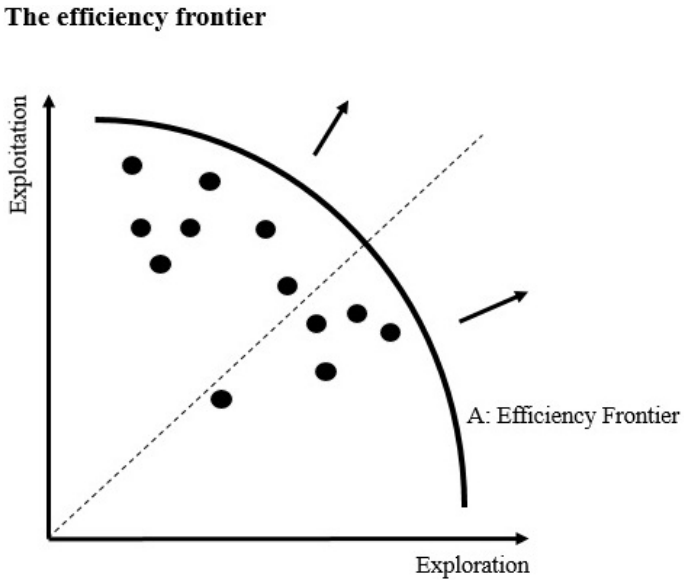

Figure 2: Different approaches to managing ambidexterity (adopted from Birkinshaw and Gupta, 2013, p. 295).

achieve ambidexterity, there are three types of choices they are facing (Birkinshaw and Gupta, 2013). The first choice is the decision on where to sit on the efficiency frontier. Birkinshaw and Gupta (2013) argue that although all positions on the efficiency frontier may be of equal value, there may be some superior positions due to certain circumstances which a firm is facing. The second choice concerns the reaching of the efficiency frontier. Managerial capability is needed to reach the efficiency frontier and some firms may be better in reconciling exploration and exploration than others (Birkinshaw and Gupta, 2013). The third choice relates to the pushing out of the efficiency frontier, meaning to move the curve up to the right. Firms which are able to push out the efficiency frontier display greater rates of ambidexterity as they are able to reconcile exploration and exploitation on a higher level through the development of new technologies and innovations (Birkinshaw and Gupta, 2013).

Pertaining to the ambidexterity – performance linkage, Birkinshaw and Gupta (2013) connect their conceptualizations shown in Figure 2 to the empirical findings of other scholars (e.g., Lin et al., 2007; Rothaermel and Alexandre, 2009). They conclude that whenever ambidexterity is measured or operationalized as the product or the sum of exploration and exploitation, the firms lying closest to the efficiency frontier are the ones who perform higher, "and therefore [...] ambidexterity matters" (p. 295). On the other hand, studies which conceptualize ambidexterity as the balance between exploration and exploitation argue that the best position is somewhere on the diagonal shown in Figure 2, meaning somewhere on the bottom-left to the top-right (Birkinshaw and Gupta, 2013).

2.4. Performance outcomes and moderators of ambidexterity

Two main domains of the prior research on ambidexterity have been the examination of the performance outcomes and moderators. Specifically, the effect of ambidexterity on the organizational performance as well as the moderating effect of environmental factors on ambidexterity have been analyzed by a variety of authors. The major outcomes of these findings will be presented in this section.

2.4.1. The ambidexterity – performance relationship

The exploration of the ambidexterity – performance relationship has been widely explored by different researchers (e.g., He and Wong, 2004; Gibson and Birkinshaw, 2004; Cao et al., 2009; Lubatkin et al., 2006). There is theoretical and empirical evidence that the coupling of exploration and exploitation, meaning to be ambidextrous, has a positive effect on a firm's performance. However, there are only few empirical findings that show the influence of the joint use of exploration and exploitation on organizational performance (He and Wong, 2004). He and Wong (2004) utter the basic assumption that explorative organizations' performance is rather variable and fluctuant due to the experimentation involved. Exploitative firms, on the other hand, can generate a more stable performance as their outcomes are more predictable and certain.

Concerning the effect of ambidexterity on a firm's performance, Raisch and Birkinshaw (2008) came to the conclusion that the relationship between ambidexterity and performance cannot yet be fully approved. Although the amount of studies concerning organizational ambidexterity is increasing, there is still little research to test the ambidexterity – performance relationship, which can partly be attributed to the lack of consensus of how exploration and exploitation should be conceptualized. Prior empirical tests in this field

(e.g., He and Wong, 2004; Gibson and Birkinshaw, 2004; Lubatkin et al., 2006) showed only limited evidence for the ambidexterity – performance relationship. According to March (1991), firms may run risk to perform poorly in either exploration or exploitation when these two activities cannot be brought into accordance with each other and thus lack balance. In addition to March's assumption that organizational performance is poor when there is no balance between exploration and exploitation, Tushman and O'Reilly (1996) suggest that the simultaneous handling of exploration and exploitation activities increases the chance of reaching greater performance than firms focusing on only one of the two activities. So, although there is still little empirical evidence on the effect of ambidexterity on performance, there is a consensus among the assumption that firms capable of simultaneously pursuing exploration and exploitation achieve superior organizational performance.

2.4.2. Environmental factors influencing ambidexterity

Environmental factors can have an influence on organizational ambidexterity. Especially the level of dynamism and competitiveness are being discussed in the literature. Jansen et al. (2006) provide an empirically supported finding for the direct effect of environmental factors on ambidexterity. The authors state that environmental dynamism and competitiveness can oblige firms to increasingly engage in exploitation and exploration and, therefore, act ambidextrously. Raisch and Birkinshaw (2008) also environmental factors have a moderating effect on the ambidexterity – performance relationship. Jansen et al. (2006) found empirical support that the pursuit of exploration is more effective in dynamic environmental conditions, whereas the pursuit of exploitation is more effective in competitive and rival environments. Raisch and Birkinshaw (2008) emphasize that becoming ambidextrous in instable environments is "more of a necessity" than a key factor which leads to greater short-term performance. Tushman and O'Reilly (1996) have found that the spatial separation of exploration and exploitation serves as one way to cope with environments of long-term stability which are interrupted by occasional periods of discontinuous change. O'Reilly and Tushman (2008) additionally state that when there are slower rates of change in the environment (incremental change), ambidexterity may be pursued sequentially, whereas in rapidly changing environments (discontinuous change) ambidexterity should be pursued simultaneously. In sum, these findings show that environmental factors influence ambidexterity. The results section of this thesis will provide greater insights into the relationship between environmental moderators and ambidexterity.

2.5. Antecedents of ambidexterity

In prior research on ambidexterity, three types of mechanisms for achieving ambidexterity are predominantly discussed: structural and contextual, and leadership-based solutions. In the prevailing literature, structural ambidexterity is mainly divided into either the spatial or the temporal separation of exploration and exploitation. Contextual ambidexterity, on the other hand, refers to the contextual integration of exploration and exploitation. Both types, apart from the subtype temporal separation, generally imply that exploration and exploitation can be pursued simultaneously. Leadership-based solutions are the third major type for achieving ambidexterity and highlight how senior managers, as key leaders in an organization, are inevitably involved in creating and fostering ambidexterity in an organization (Raisch and Birkinshaw, 2008). This section will explain structural and contextual ambidexterity, as well as leadership-based solutions for ambidexterity in more detail and describe some of the main mechanisms of how each can be achieved and managed.

2.5.1. Structural ambidexterity

Ambidexterity in an organizations can be achieved by means of structural solutions, meaning that explorative and exploitative activities are pursued in separate organizational units (Raisch and Birkinshaw, 2008). The two predominant concepts of structural solutions are the use of spatial separation and of temporal separation. Spatial separation is the more prevalent concept, having its roots in Duncan's work of 1976 who emphasizes the implementation of dual structures which involve the creation of separate units that are responsible for either explorative or exploitative activities in an organization (see Tushman and O'Reilly, 1996; Smith and Tushman, 2005). Contrary to the concept of spatial separation, temporal separation involves the use of parallel structures (e.g., McDonough and Leifer, 1983) and comprises the ability to temporally switch between exploration and exploitation.

Temporal separation implies that ambidexterity occurs sequentially, meaning that a single business units focusses on exploration the one day, and on exploitation the next day (Adler et al., 1999; McDonough and Leifer, 1983). According to the phase of innovation which an organization faces, it temporally switches and adapts its processes (O'Reilly and Tushman, 2008). Raisch and Birkinshaw (2008) highlight the primary and secondary structures (see Adler et al., 1999) which define parallel structures. Accordingly, primary structures involve mainly exploitative activities for routine tasks and for the maintenance of stability and efficiency. Secondary, or supplementary structures, involve mainly explorative activities for non-routine tasks and for innovation in order to enable efficiency and flexibility. Temporal separation allows to meet the competing demands of exploration and exploitation within the same business unit, making the coordination costs which occur in spatial separation nearly redundant (Gibson and Birkinshaw, 2004). However, according to Gibson and Birkinshaw (2004), both spatial and temporal separation still require unit managers who make decisions on how to divide up the time and groups in order to manage the trade-off between exploration and exploitation.

O'Reilly and Tushman (2008) claim that the mechanisms for temporal separation, meaning the sequential pursuit of the two activities, are quite different to the mechanisms of

spatial separation as "the challenge is transforming one internally consistent strategy and organizational alignment (e.g., a focus on efficiency or exploitation) to another" (O'Reilly and Tushman, 2008, p. 193). The authors also argue that the notion of temporal sequencing of exploration and exploitation is not always achievable due to the complexity and the pace of change in markets and technologies. Therefore, they suggest that spatial separation is the more feasible solution. According to them, spatial separation includes the simultaneous pursuit of exploration and exploitation with the implementation of separate subunits, business models, and alignments for each of the two dimensions. However, O'Reilly and Tushman (2008) also claim that "ambidexterity, in this conceptualization, entails not only separate structural subunits for exploration and exploitation but also different competencies, systems, incentives, processes and cultures – each internally aligned" (O'Reilly and Tushman, 2008, p. 193). According to Smith and Tushman (2005) this alignment across subunits can be achieved by means of a common strategic intent, shared assets, and an overarching set of values. Therefore, the top management team (TMT) has a vital function in integrating structurally separated units which are either involved in exploration or exploitation (O'Reilly and Tushman, 2008; Smith and Tushman, 2005).

Within structural ambidexterity, the way in which explorative and exploitative units are configured and organized is seemingly different from each other. On the one hand, units which are involved in exploration are said to be small, decentralized, and with loose processes (Benner and Tushman, 2003; Tushman and O'Reilly, 1996). On the other hand, units which are involved in exploration are said to be larger, more decentralized, and with tight processes (Benner and Tushman, 2003; Tushman and O'Reilly, 1996). Now, given that exploration and exploitation are separated by means of structurally differentiated units, should these different units be integrated and how can this be done? Raisch and Birkinshaw (2008) have found that some researchers emphasize the creation of "loosely coupled organizations in which the explorative units are strongly buffered against the exploitative units" (Raisch and Birkinshaw, 2008, p. 390). Contrary to this, Tushman and O'Reilly (1996), for example, argue in favor of organizational configurations that combine tight and loose coupling. This means that an ambidextrous organization can be created by tightly coupling multiple loosely coupled subunits with each other (O'Reilly and Tushman, 2004). Hereby, the explorative and the exploitative units are spatially and culturally separated and are managed by the use of different incentive systems and managerial teams. Additionally, the top management teams, which are responsible for the coordination and the development of a strong, overarching organizational culture enable the strategic integration of the different units.

2.5.2. Contextual ambidexterity

Since the focus of prior studies lied primarily on the examination of structural ambidexterity, contextual ambidexterity seems to be a neglected field of research (Raisch and Birkinshaw, 2008). This section will therefore give a definition of contextual ambidexterity and describe its mechanisms. Gibson and Birkinshaw (2004) were the first authors to investigate the antecedents, the consequences and mediating role of contextual ambidexterity in a survey of 4,195 individual in 41 business units. Therefore, their study will serve as a base for the explanation of contextual ambidexterity.

Gibson and Birkinshaw define contextual ambidexterity as "the behavioral capacity to simultaneously demonstrate alignment and adaptability across an entire business unit" (Gibson and Birkinshaw, 2004, p. 209). Alignment (exploitation) is characterized by the coherence among all different processes and activities in a business unit which are working into the direction of a common goal and is oriented towards the improvement of the performance in the short term (Gibson and Birkinshaw, 2004). Adaptability (exploration) includes the reconfiguring of these different activities in a business unit in order to quickly react to changing conditions in the task environment and is oriented towards the improvement of the performance in the long term (Gibson and Birkinshaw, 2004). The buildup of a business unit context allows individuals in an organization to undertake their own judgments concerning the division of their time when it comes to the handling of the conflicting demands of alignment and adaptability (Gibson and Birkinshaw, 2004). In other words, ambidexterity can be achieved "by building a set of processes or systems that enable and encourage individuals to make their own judgments about how to divide their time between conflicting demands for alignment and adaptability" (p. 211). Similarly, Güttel and Konlechner (2009) define contextual ambidexterity as the ability of employees to switch between explorative and exploitative activities in line with their own judgments.

Gibson and Birkinshaw (2004) argue that contextual separation is, contrary to structural separation, more efficient as it eases the adaptation of an entire business unit. When applying a structural separation model, only separate units are in charge of the adaptation to new tasks which emerge as a result of changing environmental demands. In addition to this, contextually ambidextrous business units provide a more flexible and dynamic context which enables individuals to decide which part of their time they want to spend on either alignment-oriented or adaption-oriented activities (Gibson and Birkinshaw, 2004). This means that the development of an ambidextrous context in business units promotes ambidextrous behavior on the individual level that is aligned and adaptable (Gibson and Birkinshaw, 2004). Contextual ambidexterity can therefore be defined as "an interplay of system capacities – for alignment and adaptability – that simultaneously permeate an entire business unit" (Gibson and Birkinshaw, 2004, p. 211). Güttel and Konlechner (2009) state that the main advantage of contextually ambidextrous organization designs, compared to structurally ambidextrous designs, is the facilitated and faster knowledge transfer between exploratory and exploitative activities in order to develop innovative and applicable solutions. In other words, the use of project teams facilitates the diffusion of knowl-

edge across various learning domains in contextually ambidextrous organizations, unlike structurally ambidextrous organizations where the TMT is responsible for the transfer of knowledge from exploratory to exploitative areas (Güttel and Konlechner, 2009). Therefore, the top management teams in contextual ambidextrous arrangements are not responsible for coordinating the integration across different units, but for the creation of a supportive business-unit context (Raisch and Birkinshaw, 2008). Compared to structural ambidexterity, Güttel and Konlechner (2009) state that within a contextually ambidextrous organization design, the employees possess a broad background knowledge which enhances their understanding of the demands in different domains so that "newly generated knowledge from the scientific community can be applied faster, more comprehensively and with a better fit to the client's needs in the business environment" (p. 167).

Having mentioned the main advantages of contextual ambidexterity, Gibson and Birkinshaw (2004) claim that there is a lack of understanding of how contextual ambidexterity is achieved, meaning its antecedents, and which consequences contextual ambidexterity has on business unit performance. Gibson and Birkinshaw (2004) summarize some prior recommendations for supporting contextual ambidexterity including the use of meta-routines and job-enrichment schemes, the use of leaders with complex behavioral repertoires, and the creation of a shared vision. However, as these suggestions are not comprehensive e-nough to cover all aspects of contextual ambidexterity, they additionally refer to the work of Ghoshal and Bartlett (1994; as cited in Gibson and Birkinshaw, 2004) who state that context is characterized by four behavior-framing attributes, namely discipline, stretch, support, and trust. According to Gibson and Birkinshaw (2004), organization context is related to the concepts of structural context, organization culture, and organization climate. The authors describe structural context as the development of administrative and comparatively tangible mechanisms that strengthen specific employee behaviors. Organizational culture is described as a construct which includes underlying beliefs and values of individuals in an organization. Lastly, organization climate is referred to as the presence of organizational stimuli or environmental characteristics which are assumed to influence the behaviors and attitudes of individuals. Ghoshal and Bartlett (1994; as cited in Gibson and Birkinshaw, 2004) define context as the systems, processes, and beliefs that shape behaviors on the individual level. Ideally, the context provides the possibility for individuals to decide themselves how they want to divide their time between explorative and exploitative activities. However, the creation of a supportive organization context is not enough to reach superior performance: only when the supportive context creates the capacity to achieve ambidexterity, superior performance can be reached (Gibson and Birkinshaw, 2004).

Gibson and Birkinshaw (2004) extend the concept of the four attributes by Ghoshal and Bartlett (1994) to suggest that a context is necessary for the development of a supportive environment that encourages individuals to undertake certain activities in order to achieve outcomes. More specifically, this means that the establishment of a supportive context which includes the four behavior-framing attributes, discipline, stretch, support, and trust, enables individuals to engage in alignment-oriented (exploitation) and adaptability-oriented (exploration) activities (Gibson and Birkinshaw, 2004). The results of Gibson and Birkinshaw (2004) show that the simultaneous pursuit of both activities eventuates in contextual ambidexterity which, in turn, reinforces the performance. This finding is in line with other scholars' assumptions (e.g., He and Wong, 2004) that ambidexterity has a positive effect on a firm's performance. Their hypothesis stating that the more a business unit context is characterized by an interaction of stretch, discipline, support, and trust, the higher the level of ambidexterity was also supported. Gibson and Birkinshaw (2004) have also found that there was a strong, positive correlation between alignment and adaptability, suggesting that business units are able to achieve both behaviors simultaneously.

2.5.3. Ambidextrous leadership

Senior managers, as key leaders in an organization, are inevitably involved in developing and strengthening ambidexterity in an organization (Raisch and Birkinshaw, 2008). While some studies (e.g., Gibson and Birkinshaw, 2004; Smith and Tushman, 2005) treat the top management team as a supporting factor in the implementation of ambidexterity, others (e.g., Lubatkin et al., 2006) consider leadership processes as an independent antecedent of ambidexterity. According to Raisch and Birkinshaw (2008), the latter conceptualization of leadership-based antecedents often relates to hierarchical management levels. Pertaining to this, Floyd and Lane (2000) suggest that exploration is pursued in the operating levels where managers experiment with new solutions to emerging demands, and that exploitation is pursued at the top management level where the promising solutions from the operating level are selected and deployed.

In contrast to this structural separation by the use of hierarchical levels, there is also the notion of managers who simultaneously carry out exploration and exploitation, thus following a contextual ambidexterity approach (e.g., Rosing et al., 2011). Smith (2006) refers to TMTs that shift their resources between existing products and new products to equally and simultaneously support exploration and exploitation. Similarly, Volberda et al. (2001) state that the top management manages the simultaneous pursuit of exploration and exploitation by introducing new competencies to some units while deploying existing and well-established competencies in others. In this context, Beckman (2006) also highlights the importance of the founding team composition, and especially the members' earlier company affiliations, as an antecedent of explorative and exploitative behavior. She found empirical evidence that firms in which the founding teams had both diverse and common prior company affiliations showed higher levels of ambidexterity.

Ambidextrous leaderships is the ability to promote ambidexterity among employees and to display a variety of lead-

ership behaviors, depending on the situation (Rosing et al., 2011). Rosing et al. (2011) describe ambidextrous leadership with regard to innovation. The authors state that there are two processes involved in innovation, namely creativity (exploration) and implementation (exploitation). As the two activities, exploration and exploitation, are very different to each other and sometimes even opposing, the challenge is to flexibly switch between them (Rosing et al., 2011). This implies that "teams involved in innovation need to show exploration and exploitation in an unpredictably alternating sequence" (Rosing et al., 2011, p. 966). This assumption is rooted in the contextual ambidexterity approach, meaning that teams or business units are not structurally separated in order to pursue either exploration or exploitation but that they inevitable engage in both domains. Also, with regard to ambidextrous leadership, it is supposed that exploration and exploitation are not mutually exclusive, but can occur simultaneously. Rosing et al. therefore propose that "it is necessary not only to be able to balance exploration and exploitation, but to be able to integrate exploration and exploitation and flexibly switch between both as the situation requires" (Rosing et al., 2011, p. 966).

According to Rosing et al. (2011), ambidextrous leadership comprises three elements, namely opening leader behavior, closing leader behaviors, and the temporal flexibility to switch between both depending on the situation. Opening leader behavior is related to fostering exploration among followers. It includes the breaking up of established routines as well as the thinking in new directions. Leaders who display an "open" behavior give space for independent thinking and acting, encourage their followers to experiment and to approach things differently, and promote (risky) efforts with the aim of changing existing routines. Closing leader behavior is related to fostering exploitation among followers. It includes rationalization in order to reduce variance and to support the implementation of routines. Leaders who display a "closed" behavior set guidelines, are in charge of correcting certain actions, and supervise the achievement of goals.

In addition to this, Rosing et al. (2011) explain that both opening and closing leader behaviors can be performed actively or passively. Actively means that leaders themselves pursue explorative or exploitative activities. Passively, on the other hand, refers to the degree of promoting explorative or exploitative behaviors among the leaders' followers and giving them room for working independently. When a leader displays an open leader behavior, this may for example manifest itself in actively introducing new ideas and in passively encouraging error learning. When a leader displays a closed leader behavior, this may manifest itself in actively taking corrective actions or structuring tasks and in passively controlling the adherence to goals. In conclusion, Rosing et al. (2011) highlight the importance of being able to have a repertoire of both closing and opening leader behavior although these two are very different from each other. Lubatkin et al. (2006) similarly found that top management team behavioral integration positively correlated with firm performance and that this correlation was mediated by ambidexterity. The term "behavioral integration" refers to the degree of a top management team's wholeness and unity of effort and is determined by the level of the team's collaborative behavior, the quantity and quality of information exchanged, and the degree of joint decision making (Lubatkin et al., 2006).

With regard to ambidextrous leadership, Rosing et al. (2011) also mention that there might be multiple leaders in a team who are responsible for promoting exploration and exploitation among their followers. However, this requires coordination with regard to when opening or closing behaviors need to be displayed by the different leaders within a team. Additionally, Rosing et al. (2011) state that a team's culture or climate may have an impact on leadership behaviors. For example, when a team ensures controlled goal achievement and high standards of performance, then a transformational leadership style leads to a high innovative performance (Rosing et al., 2011). Tushman et al. (2011) claim that great leaders manage the tensions between new innovations (exploration) and core products (exploration) by developing an overarching vision, by holding tension at the top meaning that decisions need to be made at the senior-executive level, and by embracing inconsistency in different business units.

2.6. Conclusions

The examination of the conceptual background served as a basis for the appraisal of the mechanisms through which ambidexterity can be achieved and to show how exploration and exploration can be reconciled in practice. However, the findings of prior research on ambidexterity have not always been congruent and sometimes show contradictory results. First of all, the concept of ambidexterity has been examined through a variety of literature streams which leads to a multitude of different approaches on how to tackle this topic. The different literature streams which were explained included organizational learning, technological innovation, organizational adaptation, strategic management, and organizational design.

Secondly, there is the question of how ambidexterity, and more specifically exploration and exploitation, should be conceptualized. Accordingly, there are two different possibilities: continuity and orthogonality. While conceptualizing exploration and exploitation as two ends of continuum, thus being mutually exclusive, orthogonality implies that high levels of either exploration or exploitation in one domain may simultaneously be present with high levels of exploration or exploitation in the other domain. Building on these two different conceptualizations the question arises of whether to pursue a punctuated equilibrium approach or ambidexterity. On the one hand, punctuated equilibrium includes a sequential or cyclical pursuit of exploration and exploitation. On the other hand, ambidexterity involves the simultaneous pursuit of the two domains. According to this, punctuated equilibrium seems to be the better solution when exploration and exploitation are conceptualized as two ends of a continuum. Ambidexterity is the better solution when exploration and exploitation are orthogonal. Due to the fact that the

punctuated equilibrium approach presumes that exploration and exploitation are pursued sequentially and not simultaneously, practical implications regarding this conceptualization will be neglected in the later parts of this thesis.

Thirdly, there are two main types of ambidexterity: structural and contextual ambidexterity. Structural ambidexterity works by spatially separating business units which enables specialization in both exploration and exploitation. Contextual ambidexterity involves the contextual integration of individuals or subsystems who simultaneously pursue exploration and exploitation. While the role of the top management team in structural ambidexterity is the cross-linkage and communication across the different specialized business units, the top management team in contextual ambidexterity is responsible for the creation of a contextual framework (e.g. a context characterized by a combination of stretch, discipline, support, and trust). Apart from structural and contextual solutions of achieving and managing ambidexterity, leadership-based solutions are a third mechanism in order to strengthen ambidexterity in an organization. Leadership-based solutions mainly refer to certain characteristics of leaders or top management teams which should facilitate ambidexterity. Ambidextrous leadership can therefore be regarded as another antecedent of ambidexterity.

Fourthly, the level of analysis for ambidexterity is still ambiguously treated in the literature. Research so far mainly considers ambidexterity on the individual, the business unit, and the organizational level. Concerning the studying of the ambidexterity-performance linkage, most researchers analyzed this relationship on the organizational level (e.g., He and Wong, 2004). The organizational level of analysis is, together with the business unit level, generally used to examine structural ambidexterity (e.g., Tushman and O'Reilly, 1996; Benner and Tushman, 2003; Lubatkin et al., 2006; Jansen et al., 2008, 2009). Studies focusing on contextual ambidexterity mainly analyzed the individual or the business unit levels (e.g., Gibson and Birkinshaw, 2004; Güttel and Konlechner, 2009). According to these findings, the level of analysis is related to the type of ambidexterity that is examined.

Lastly, the discussion about the rigor-relevance debate delivered important findings concerning the core difficulties related to the research question. As this thesis aims at providing practical implications for managers concerning the concept of ambidexterity, the translation of scientific findings into practical implications can be regarded as a main challenge. The following sections of this thesis will thus try to elaborate an appropriate method for bridging the rigor-relevance gap concerning the topic of ambidexterity and its practical implications. The aim is to derive implications for managers from academic as well as practitioner literature for achieving ambidexterity in practice. Contrary to prior studies, which strictly separated structural, contextual, or leadership-based solutions, this thesis sees management as an overarching element with which ambidexterity can be achieved. In other words, the top management team has an all-embracing role in recognizing the need for ambidexterity in an organization and in implementing appropriate actions for the coupling of exploration and exploitation. This largely untreated area of ambidexterity research, thus, requires further examination and analysis. The aim of this thesis is, therefore, to provide an overview of the major findings regarding the practical implications for managers which can be found in academic and practitioner literature for achieving ambidexterity in practice. The next section will explain the methodological proceeding with regard to this purpose.

3. Research strategy and method

In order to define a relevant research question the existing literature with regard to the thesis' topic was examined. This was done in the above section 2, the analysis of the conceptual background. The aim of this literature study was to get an overview of the current state of research in the field in order to enable the specification of the research question with the intention of enlarging the prevailing knowledge base (Tranfield et al., 2003). The scoping of the literature concerning the topic of ambidexterity revealed that there are multiple organizational levels in which there are different processes and mechanisms for coupling exploration and exploitation (i.e. structural, contextual, and leadership-based solutions). The management has an overarching function in implementing these mechanisms. The proposed research question which will be treated in the following parts of this thesis is the following:

> Which practical implications for managers can be found in academic and practitioner literature for achieving ambidexterity in practice?

In order to answer this research question and thus to fill the research gap, a systematic literature review offers an applicable methodology for the assessment of the current knowledge concerning ambidexterity. Specifically, it will be examined which practical implications can be derived from academic literature as well as practitioner literature regarding the concept of ambidexterity. The results obtained from the academic and the practitioner literature will be analyzed and interpreted separately. This should allow a differentiated evaluation of the findings concerning the practical implications of the ambidexterity concept. While the results obtained from the practitioner literature might be more practice-oriented, those obtained from the academic literature are supposed to be more general and scientific. This assumption is leaned on the debate of the rigor-relevance gap in management research. As already mentioned earlier, the rigor-relevance discussion is about the question of whether scientific research can be connected to practice (and vice versa) or not (see Kieser and Leiner, 2009; Hodgkinson and Rousseau, 2009). Therefore, if this thesis succeeds in deriving relevant practical implications regarding the concept of ambidexterity, the rigor-relevance gap can be considered bridgeable.

3.1. Why a systematic review?

The topic "Practical Implications of the Ambidexterity Concept" will use a systematic review method in order to

clarify the question of the extent to which practical implications for managers or practitioners can be found in the existing literature on ambidexterity. The work of Tranfield et al. (2003) will be used to explain how evidence-informed research can be conducted by means of a systematic review. In the field of management, the endeavor of conducting a systematic review is liable to several problems, such as the heterogeneity of studies and the adjunctive amount of different research questions addressed in these studies, making it difficult to synthesize and to associate the findings with each other (Tranfield et al., 2003). A systematic review should identify key scientific findings with regard to the research question. Tranfield et al. (2003) argue that a meta-analysis, contrary to a systematic review, uses a statistical procedure to synthesize the findings from a broad range of studies which is not always appropriate in the field of management. They reason this by stating that in management research only few studies address the same research question or measure certain phenomena in the same way, making it difficult to use a meta-analytic approach. A systematic literature review, however, allows a transparent, reproducible, comprehensive, and unbiased search.

A systematic review should furthermore help to solve the problem of bridging rigor and relevance in the management field. Hodgkinson et al. (2001) classify applied social sciences into four categories: "Popularist Science", "Pedantic Science", "Puerile Science", and "Pragmatic Science" (see figure 3). "Popularist Science" is classified as research that is high on practical relevance, but low on methodological rigor. "Pedantic Science", on the other hand, is research that is high on rigor but low on relevance. "Puerile Science" is research that is neither relevant nor rigorous. "Pragmatic Science" is the only research which displays both rigor and relevance on a high level. With regard to this, Tranfield et al. (2013) state that a "systematic review can be argued to lie at the heart of a 'pragmatic' management research, which aims to serve both academic and practitioner literature" (p. 220).

The main problem regarding the rigor-relevance discussion is to investigate whether there are practical implications concerning the ambidexterity concept or not. If no practical implications can be derived from the prevailing literature (i.e. when the theoretical examination of the topic remains dominant), the existence of the rigor-relevance gap concerning the topic of ambidexterity can be approved. The systematic review of academic journals and practitioner reviews should help to investigate the rigor-relevance debate concerning ambidexterity. The assumption hereby is to possibly deduct more concrete implications for managers from practitioner literature (i.e., in Harvard Business Review, MIT Sloan Management Review, and California Management Review) and more abstract and general instructions from academic journals.

The following sections will describe the different stages of conducting a systematic literature review. The work of Tranfield et al. (2003) will provide the theoretical basis. The authors have established three main stages: planning the review, conducting the review, and reporting and dissemination. Planning the review includes the conduct of scoping studies and the setup of a review protocol which documents information on the various articles. The phase of conducting a review includes the definition of a search strategy, the creation of data extraction forms which contain specific information about the selected studies, and the synthesis of the data. For the data synthesis, the process of categorization and the qualitative (and quantitative) analysis of the content are described with reference to the work of Mayring (2015). Lastly, the phase of reporting and dissemination will concern the conclusions and recommendations for practitioners derived from the systematic literature review.

3.2. Planning the review

The first stage of the systematic review process is to plan the review. This was done by conducting scoping studies with the aim of getting an overview of the relevance and the size of the existing literature in order to narrow down the topic. The section about the conceptual background above considered different perspectives of the ambidexterity concept and showed how this topic was approached by other researchers. The analysis of the conceptual background not only gave an overview of the theoretical findings, but also included the practical and methodological history in the field of ambidexterity. Tranfield et al. (2003) note that a review protocol should be set up in form of a plan which documents the single steps taken. This plan ensures objectivity and also serves as a point of reference when writing the thesis. The recording of the results was done in an excel file which includes columns presenting the formal aspects including the name of the author(s), the year of publication, the title of the paper, the name of the publishing journal, the research focus of the paper, the theoretical lens, its approach to ambidexterity (i.e. structural or contextual), the level of analysis (i.e. individual, group, organization), the methodology used, and the key findings.

3.3. Conducting the review

The second stage concerns the actual conduct of the systematic review which includes the systematic search with keywords and search terms, the selection of studies based on the scoping study conducted for the conceptualization of the theoretical background of ambidexterity, and the synthesis of the data (Tranfield et al., 2003). Again, the processes of searching should be reported in a review protocol including a full listing of all articles in order to ensure the later replication of the study (Tranfield et al., 2003). In order to allow a differentiated analysis of the results obtained from either the academic or the practitioner literature, these findings will be evaluated separately. Two independent searches were conducted for the selection of practical implications concerning ambidexterity from academic and practitioner literature. The first search included the selection of the 40 most cited articles from academic journals treating ambidexterity. The second search concerned the identification of studies in practitioner literature which treated the topic of ambidexterity.

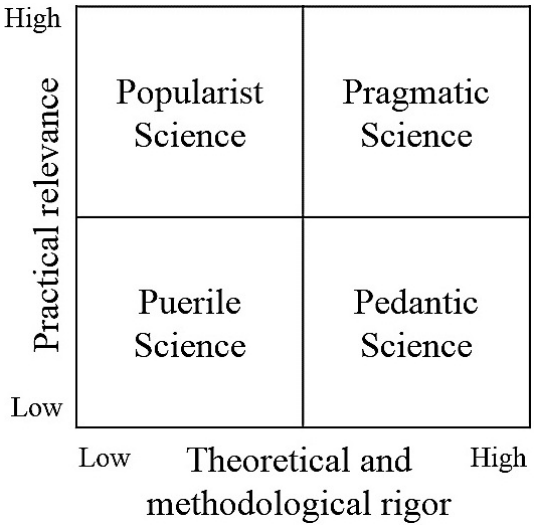

Figure 3: Characterization of applied social science (adopted from Hodgkinson et al., 2001, p. 42).

3.3.1. Search strategy: selection of studies and data extraction

For the answering of the research question, the 40 most cited articles concerning the topic of ambidexterity were selected for the later evaluation. The search was restricted to the 40 most cited articles from academic journals only in order to represent the most influential studies concerning the concept of ambidexterity. The database Web of Science was used for this search. The criteria of search were the following: 'ambidext*' (restricted to topic) was used as the search term; the selected time span reached from 1996 to 2014; the research categories were set to 'Management' and 'Business'; the results were sorted from the amount of times cited from highest to lowest. With this search, 414 results (on 05.06.2015) were obtained. The search criteria were documented in a review protocol. According to Tranfield et al. (2003) only those studies that meet all the inclusion criteria should be incorporated into the review. Apart from the time span (articles published from 1996 to 2014) and a VHB-JOURQUAL3 ranking of C or better, the only other inclusion criterion required that the selected article treated ambidexterity in some way.

A second search was conducted using EBSCOhost in order to find articles published in practitioner literature which were possibly neglected in the search of the 40 most cited articles. Again, the search term 'ambidext*' was used to show results which include this term in the title. Furthermore, the search was restricted to articles published in three practitioner reviews, namely in California Management Review, Harvard Business Review, and MIT Sloan Management Review. This search led to six results. Three of the studies overlapped with the results from the search of the 40 most cited articles. This was compensated by not including these results in the 40 most cited articles. Due to the limited number of results obtained in the search for studies in practitioner literature, no further inclusion or exclusion criteria were specified.

Tranfield et al. (2003) note that the selected studies should be evaluated according to their internal validity and the degree to which its design, method, and analysis have minimized errors and biases. Due to the fact that the first search, meaning the search for the 40 most cited articles referring to the topic of ambidexterity, was limited to those studies who showed the most frequent citations by other authors, the evaluation of the studies' internal validity was neglected. It was assumed that the most commonly cited articles equally represent the most influential ones. Concerning the second search, the search for studies in practitioner literature, the results were limited in number, making the appraisal of the articles' internal validity dispensable.

A further step of conducting a systematic review is the process of data extraction and monitoring (Tranfield et al., 2003). Data-extraction forms aim at reducing human error and bias and include general information, i.e. the name(s) of the author(s), the publishing year, and the name of the journal), a brief synthesis of the key results of the paper, and the number of citations of each article. Table 1 and 2 present the results of the selected articles from the academic and the practitioner literature and should also serve as a data-repository for the later analysis of the data.

3.3.2. Analysis of the data: qualitative content analysis and categorization

According to Tranfield et al. (2003) "research synthesis are methods for summarizing, integrating, and, where possible, cumulating the findings of different studies on a topic or research question" (p. 217). The result of this data synthesis where the different findings on a specific topic are cumulated, is to generate a deeper understanding and to achieve a level of conceptual and theoretical development. In this thesis, the aim is to derive practical implication for managers concerning the topic of ambidexterity. The analysis of the content requires a categorization in depth and width, which means that the content from the studies needs to be filtered out analytically with the help of categories. The analysis of quali-

Table 1: Data extraction form of the 40 most cited articles from academic literature

Author(s) & Year	Methodology	Key issues & findings	Nb. of citations
Voss et al. (2008)	Survey of non-profit professional theaters in US	The authors examine how slack resources (financial, operational, customer relational, and human resource) interact with environmental threat appraisal to influence exploration and exploitation. Operational and HR slack are positively related to exploitation. When the environment is perceived as more threatening, financial and customer relational slack are positively (or less negatively) related to exploration.	101
Uotila et al. (2009)	Survey of 279 S&P 500 manufacturing firms	The study set out to empirically test the relationship between a firm's exploration and exploitation and its market-based performance. There is a curvilinear relationship between the relative amount of exploration and financial performance. Also, this relationship was found to be more pronounced in R&D intensive industries.	100
Tiwana (2008)	Survey of 42 innovation-seeking project alliances	The author examines the tensions and complementaries between bridging ties and strong ties in influencing ambidexterity of innovation alliances. A high level of knowledge integration among the collaborator in an alliance is positively related to alliance ambidexterity. Strong ties complement bridging ties in facilitating knowledge integration.	89
Taylor and Helfat (2009)	Case study of technological transitions at IBM	The authors conceptualize organizational linkages between the new technology and existing assets during transitions.	55
Smith and Tushman (2005)	Theoretical study	The authors examined decision processes in top management teams to manage the strategic and information processing contradictions associated with balancing exploration and exploitation. Achieving ambidexterity requires a paradoxical cognition among senior managers which can be cultivated by following either a leader-centric (team interactions, supportive coaching) or a team-centric (shared mental models) approach.	256
Smith and Lewis (2011)	Survey of studies focusing on organizational paradox	The authors present a dynamic equilibrium model of organizing, which depicts how cyclical responses to paradoxical tensions enable sustainability—peak performance in the present that enables success in the future. This review and the model provide the foundation of a theory of paradox.	111
Simsek et al. (2009)	Conceptual paper	The authors identify four archetypes of ambidexterity, using a temporal (simultaneous vs. Sequential) and a structural (independent vs. Interdependent) dimensions. The identified types of ambidexterity include harmonic, cyclical, partitional, and reciprocal ambidexterity. The theoretical grounding, the antecedents, and the outcomes of each type are described.	53
Simsek (2009)	Conceptual paper	The authors review previous research on the conceptualization, antecedents, and consequences of ambidexterity. They investigate the relationship between network centrality/diversity and organizational ambidexterity; also the moderating effects of dual structures, behavioral context, TMT behavioral integration, environmental dynamism and complexity are considered.	77

(Continued)

Table 1—continued

Author(s) & Year	Methodology	Key issues & findings	Nb. of citations
Sidhu et al. (2007)	Cross-sectional survey in manufacturing industries	The authors conceptualize exploration and exploitation in terms of a nonlocal-local search continuum in three-dimensional supply, demand, and geographic space. Boundary spanning supply-side search is found to be positively associated with innovation in more-dynamic environments typical of the entrepreneurial regime phase of technology evolution. Boundary-spanning demand-side search is found to be favorably associated with innovation in less-dynamic environments. Spatial boundary-spanning search seems to contribute to innovation in more- as well as less-dynamic environments.	75
Sheremata (2000)	Theoretical paper	The author finds that ambidexterity in new product development projects increases the likelihood that project teams will find high-quality solutions to problems quickly and efficiently while maintaining balance among their goals. Ambidexterity is, in turn, positively related to the attainment of development schedule, cost, and product quality goals.	137
Rothaermel and Alexandre (2009)	Survey of the manufacturing sector in the US	The authors examine four possible combinations of exploration and exploitation, based on the technological boundary (new or known knowledge) and the organizational boundary (internal or external sourcing). They found that an inverted U-shaped relationship exists between a firm's total technology sourcing mix (of known and new technology) and its performance; and a firm's absorptive capacity moderates the inverted U-shaped relationship between a firm's total technology sourcing mix (of known and new technology) and firm performance in such a fashion that the positive effect of ambidexterity in technology sourcing on firm performance is stronger when the firm possesses higher levels of absorptive capacity.	104
Raisch et al. (2009)	Theoretical study	The authors explored four fundamental tensions related to organizational ambidexterity, including differentiation versus integration, individual versus organizational, static versus dynamic, and internal versus external. Research on organizational ambidexterity shows that some individuals, groups, and organizations are successful in the long run. The authors provide important insights into the strategies, structures, and processes that allow them to balance and harmonize seemingly contradictory requirements.	226
Raisch and Birkinshaw (2008)	Review article	The authors show that ambidexterity spans various (disconnected) research fields and highlight research done on antecedents (structural, contextual, leadership-based), the ambidexterity-performance linkage, environmental factors and other moderators.	262
O'Reilly and Tushman (2008)	Theory paper	The authors identify a set of propositions that suggest how ambidexterity acts as a dynamic capability. They suggest that efficiency and innovation need not be strategic tradeoffs and highlight the substantive role of senior teams in building dynamic capabilities.	149

(Continued)

Table 1—continued

Author(s) & Year	Methodology	Key issues & findings	Nb. of citations
Mom et al. (2009)	Survey of 716 managers in 5 large firms	Findings regarding the formal structural mechanisms indicate that a manager's decision-making authority positively relates to this manager's ambidexterity, whereas formalization of a manager's tasks has no significant relationship with this manager's ambidexterity. Regarding the personal coordination mechanisms, findings indicate that both the participation of a manager in cross-functional interfaces and the connectedness of a manager to other organization members positively relate to this manager's ambidexterity. Furthermore, results show positive interaction effects between the formal structural and personal coordination mechanisms on managers' ambidexterity.	69
Luo and Rui (2009)	Survey of multiple case studies from Chinese firms	The authors conceptualize ambidexterity as a multidimensional term comprising co-evolution, co-competence, co-opetition, and co-orientation. They show how emerging market multinational enterprises (EM MNEs) use these four dimensions to compete against other firms in the market.	50
Lubatkin et al. (2006)	Survey of CEOs and TMT members from 139 SMEs	TMT behavioral integration is positively associated with the ambidextrous orientation of SMEs. Behaviorally integrated TMTs (collaboration, joint decision-making, information exchange) are better able to handle the informational contradictions and conflicts associated with ambidexterity. Ambidextrous orientation is positively related to relative firm performance (growth and profitability) among SMEs.	221
Lin et al. (2007)	Archival study of 5 US industries and computer simulation model	The authors examine the boundary conditions under which ambidexterity improves firm performance. Firm size, environmental uncertainty, and network centrality weaken the effect of ambidexterity on performance; also, the impact of ambidexterity on performance is stronger in the early years of network formation. A high degree of structural holes in Interfirm networks negatively moderates the impact of alliance ambidexterity on firm performance.	68
Lichtenthaler and Lichtenthaler (2009)	Theoretical study	The authors identify six 'knowledge capacities' as a firm's critical capabilities of managing internal and external knowledge in open innovation processes: inventive, absorptive, transformative, connective, innovative, and desorptive capacity. A thorough analysis of the knowledge capacities and knowledge management capacity reveals which capacities are well developed and where a firm has deficits.	109
Kaplan and Henderson (2005)	Theoretical study	The authors use the problems experienced by established firms attempting to create new businesses to focus attention on the forces that shape and constrain the development of new incentive systems.	61
Kang and Snell (2009)	Theoretical study	The authors find that refined interpolation (combination of specialist human capital, cooperative social capital, and organic organizational capital) and disciplined extrapolation (generalist HC, entrepreneurial SC, and mechanistic OC) facilitate ambidextrous learning.	76

(Continued)

Table 1—continued

Author(s) & Year	Methodology	Key issues & findings	Nb. of citations
Jansen et al. (2006)	Survey of 283 organizational units of a large European financial services firm	Results indicate that centralization negatively affects exploratory innovation, whereas formalization positively influences exploitative innovation. Interestingly, connectedness within units appears to be an important antecedent of both exploratory and exploitative innovation. Furthermore, the findings reveal that pursuing exploratory innovation is more effective in dynamic environments, whereas pursuing exploitative innovation is more beneficial to a unit's financial performance in more competitive environments.	324
Jansen et al. (2009)	Random company sample (4.000 firms, 230 replied)	The authors find that the direct effect of structural differentiation on ambidexterity operates through informal senior team (i.e., senior team social integration) and formal organizational (i.e., cross-functional interfaces) integration mechanisms.	109
Jansen et al. (2008)	Survey of 305 senior team members and 89 executive directors at Dutch autonomous branches of a large European financial services firm	The authors examine the relationship between senior team attributes and organizational ambidexterity. Senior team shared vision and senior team contingency rewards are positively related to ambidexterity. Transformational leadership behavior positively moderates the impact of senior team social integration and negatively moderates the effect of contingency rewards on ambidexterity.	71
Im and Rai (2008)	Online survey of customers and vendors of a company in the US logistics industry	The authors examine the impact of exploitative and explorative knowledge sharing in interorganizational relationships on relationship quality. Long-term relationships with a simultaneous emphasis on exploitative and explorative knowledge sharing experienced lower intra-group variance-to-mean performance than those that focus primarily on explorative knowledge sharing.	70
Hotho and Champion (2011)	Case study of a SME in the computer games industry	The authors examine changing people management practices as the case company undergoes industry-typical strategic change to embark on explorative innovation and it seeks to argue that maintaining an organizational context conducive to innovation over time risks turning into a contest between management and employees, as both parties interpret organizational pressures from their different perspectives.	81
Helfat and Winter (2011)	Theoretical study	The authors state that it is impossible to draw a bright line between dynamic and operational capabilities because: 1) change is always occurring to at least some extent; 2) we cannot distinguish dynamic from operational capabilities based on whether they support what is perceived as radical versus non-radical change, or new versus existing businesses; and 3) some capabilities can be used for both operational and dynamic purposes.	50
He and Wong (2004)	Survey of 206 manufacturing firms in Singapore and Malaysia	The authors find that the interaction between explorative and exploitative innovation strategies is positively related to sales growth rate and that the relative imbalance between the two is negatively related to sales growth rate.	525

(Continued)

Table 1—continued

Author(s) & Year	Methodology	Key issues & findings	Nb. of citations
Gupta et al. (2006)	Conceptual paper	The authors examine the meaning of exploration and exploitation, whether they are two poles on a continuum or orthogonal and how a balance between these two should be achieved - via ambidexterity or punctuated equilibrium.	469
Gulati and Puranam (2009)	Survey of the networking company Cisco	The authors explain how inconsistencies between formal and informal organization arising from reorganization can help create ambidextrous organizations. Compensatory fit is when, under some conditions, the informal organization can compensate for the formal organization by motivating a distinct but valuable form of employee behavior that the formal organization does not emphasize, and vice versa.	49
Gibson and Birkinshaw (2004)	Survey of 4,195 employees in 41 business units of 10 multinational firms	The authors investigated the concept of contextual ambidexterity on business unit level. They found that a context characterized by a combination of discipline, support, stretch, and trust facilitates contextual ambidexterity which subsequently leads to superior performance. Also, contextual ambidexterity mediates the relationship between these four contextual features and performance.	506
Eisenhardt et al. (2010)	Theoretical study	The authors state that, regarding structure, balancing efficiency and flexibility comes, counterintuitively, from unbalancing in favor of flexibility. Regarding environment, environmental dynamism is a multidimensional construct that can be unpacked into dimensions such as ambiguity and unpredictability. Regarding cognition, effective leaders can manage the cognitive contradiction inherent in balancing efficiency and flexibility by relying on higher-order thinking and expertise.	59
Dess and Lumpkin (2005)	Theoretical study	Firms that want to engage in successful corporate entrepreneurship need to have an entrepreneurial orientation. There are five dimensions of entrepreneurial orientation: autonomy, innovativeness, proactiveness, competitive aggressiveness, and risk-taking.	79
O'Connor and DeMartino (2006)	Longitudinal study of 12 large multinational firms	The authors investigate how large organizations can foster radical innovations internally and caution that the structural separation approach may be insufficient to develop longer-term organizational capability. They identify that a model of discovery–incubation–acceleration is beneficial in supporting commercialization.	67
Cao et al. (2009)	Survey in 3 high-tech parks in China	The authors describe two dimensions of ambidexterity: the balance and the combined dimension. Small firms with little resources benefit from a trade-off, a balance, between exploration and exploitation. Large firms operating in environments which provide sufficient resources, benefit from simultaneously combining high levels of exploration and exploitation respectively.	118
Benner and Tushman (2003)	Theoretical study	Ambidextrous organizations composed of multiple tightly coupled subunits are loosely coupled with each other. Strategic integration of the subsystems is facilitated by (heterogeneous) senior teams who promote common aspirations.	695

(Continued)

Table 1—continued

Author(s) & Year	Methodology	Key issues & findings	Nb. of citations
Beckman (2006)	Survey of 170 US high-tech firms; Multi-method design	The founding team's prior company affiliations (common vs. diverse) affect the pursuit of exploratory and exploitative strategies. The simultaneous pursuit of exploration and exploitation requires TMTs to draw on member's common and unique affiliations.	116
Andriopoulos and Lewis (2009)	Comparative case study of 5 ambidextrous firms leading the product design industry	The authors examine how executives embrace the paradoxes of strategic intent, customer orientation, and personal drivers through a combination of integration (contextual) and differentiation (structural). Three factors interact to reinforce and sustain organizational ambidexterity: a multilevel approach, complementary tactics, and learning synergies.	136
Ambos et al. (2008)	Survey of 207 academic research projects	The authors examine how the capacity of two activities (academic rigor and commercialization) can simultaneously be developed at organizational and individual level.	52
Adler et al. (1999)	Case study of the Toyota Production System	Four kinds of organizational mechanisms that can help shift the tradeoff between efficiency and flexibility: meta-routines, enrichment, switching, and partitioning. Key features of context include trust and training.	316

tative data includes four main steps: the determination and examination of the source material, the process of inductive category-building, coding and revision of the categories, and the analysis and interpretation of the category system. The following instructions for the systematical proceeding of the content analysis, and especially for the process of categorization, are based on the work of Mayring (2015). According to Mayring (2015), the category system is the main instrument of a content analysis and should correspond to the general quality criteria, namely reliability (i.e. reproducibility and accuracy) and validity.

Determination and examination of the source material

The determination of the source material is the first phase in conducting a qualitative content analysis (Mayring, 2015). First of all, the material which will be used in the content analysis needs to be defined. For this thesis, the material includes the 40 most cited articles in academic journals and six studies from practitioner literature concerning the topic of ambidexterity. As a second step, it needs to be determined how and by whom the source material was produced. The material, meaning the 46 studies in total, were selected according to the search strategy mentioned above. Thirdly, it is essential to determine the characteristics of the material and in which form it is available. In this thesis, the material is available in written form and was stored both electronically and in hard copy. The electronic administration of the literature was done with the help of 'Citavi'. The second phase of a qualitative analysis of the content is to determine what one wants to analyze from the source material (Mayring, 2015). In this literature work, the research question serves as the basis for the interpretation of the material. The point of origin is therefore the investigation of practical implications concerning the concept of ambidexterity which can be derived from academic and practitioner literature.

The examination of the material was done by screening the literature and highlighting practical implications which were relevant for the topic of ambidexterity in the text. The results of the examination of the content were again documented in an excel file. The information included in this document were general information (i.e. name(s) of the author(s) and the publication year), and the text modules referring to the practical implications of the ambidexterity concept. Additionally, a description of the text modules was also included whenever necessary. The wording of the relevant text modules (and the descriptions) was directly adopted from the original text. In order to ensure that the text modules can easily be set into context, the excel file includes a page number of each. The results of this documentation can be found in appendices A, B, and C. The adoption of the original wording should allow to derive categories in which the different text modules can be categorized. This will be explained in more detail in the next section.

Inductive category-building

The third phase of a qualitative content analysis concerns the selection of concrete techniques of analysis (Mayring, 2015). For this thesis, the process of inductive category-building was considered most appropriate. The question that arises is whether a deductive or an inductive definition of the categories is more appropriate for the analysis of the content? For this thesis, a mixture of both inductive and deduc-

Table 2: Data extraction form of the selected articles from practitioner literature

Author(s) & Year	Methodology	Key issues & findings	Nb. of citations
Tushman et al. (2011)	In-depth study of 12 top-management teams in major companies	The authors state that firms thrive when senior managers embrace the tension between new and old and foster a state of constant creative conflict at the top. There are three principles to achieving ambidexterity: develop an overarching identity, hold tension at the top, and embrace inconsistency.	13
Tushman and O'Reilly (1996)	Field research in multinational firms	The authors state that as organizations go through periods of evolutionary or revolutionary change, they need to align their competencies, strategies, structures, cultures, and leadership skills. Structural (autonomous business units), cultural (loose-tight culture), and management (ambidextrous managers, coherent vision) factors facilitate the simultaneous pursuit of incremental and discontinuous innovation and change.	637
O'Reilly et al. (2009)	Case study	The authors propose a theoretical explanation of how organizational adaptation (variation, selection, retention) can occur and provided a qualitative illustration for how this might work in practice (IBM). They conclude that a combination of a clear strategic intent, guaranteed funding, senior-level sponsorship, entrepreneurial leaders, and an aligned organization were required for the venture to succeed.	31
O'Reilly and Tushman (2011)	Semi-structured interviews with senior managers in 15 firms	The authors show that the most-successful ambidextrous designs had leaders who developed a clear vision and common identity, built senior teams that were committed to the ambidextrous strategy and were incented to both explore and exploit, employed distinct and aligned subunits to focus on either exploration or exploitation, and built teams that could deal with the resource allocations and conflicts associated with exploration and exploitation.	35
O'Reilly and Tushman (2004)	Theoretical study	The authors examined the characteristics of firms that have been successful at balancing exploration and exploitation by creating organizationally distinct units that are tightly integrated at the senior executive level. Ambidextrous organizations need ambidextrous senior teams (executives with the ability to understand the needs of different businesses, articulate a clear and compelling vision, and demonstrate commitment to ambidexterity.	337
Birkinshaw and Gibson (2004)	Survey of 4,195 individuals across 41 business units in 10 multinational companies	The authors identified four ambidextrous behaviors in individuals and five pathways for executives who want to build an ambidextrous organization. There are two dimensions of organizational context: performance management (stretch and discipline) and social support (support and trust).	102

tive building of categories was used. Deductive means that specific categories have already been defined while screening the literature or have been derived from the current state of research (Mayring, 2015). As mentioned earlier, the management has a vital and overarching function in implementing mechanisms for the achievement of ambidexterity. Therefore, three main categories were (deductively) defined as follows: measures at the top management team level, measures affecting the implementation within the organization, and moderators and other external factors. Measures at the TMT level include all actions that are implemented directly at this level, such as paradoxical cognition, strategy-making, or TMT-constellations. This category also includes personal characteristics of leaders or leadership styles which should facilitate ambidexterity. The second category, meaning the mechanisms which affect the actual implementation of am-

bidexterity within the remaining organization, include structural and contextual arrangements, as well as human resource practices. The third category refers to moderators and other external factors affecting the achievement of ambidexterity, which are more or less out of scope for decision-making and which can only marginally be controlled by managers. Subcategories which were classified into the corresponding main category were derived inductively. Inductive means that the categories do not refer to any of the earlier theory of the topic but are derived directly from the material and specific text modules (Mayring, 2015). Any text module that referred to a practical implication of ambidexterity fulfilled the criterion of selection and was assigned to a main and a sub-category for the later analysis.

Coding and revision of the category system

Coding refers to the attribution of different text modules to a defined category. Mayring (2015) explains the process of coding as follows. As soon as the criterion of selection, taking account of the level of abstraction, has been fulfilled by one text module in the material, the first category can be defined by more or less adopting its wording when formulating a term or sentence to label the new category. When a second text module fulfills the criterion of selection it can be decided whether the text module fits the already existing category (subsumption) or whether a new category needs to be defined. After the sighting of the material from 10-50% it needs to be decided whether the criterion of selection has been well defined and aids in building categories which help to answer the research question. If this is not the case, the analysis of the material needs to be repeated from the very beginning. Otherwise, the process can be continued.

Analysis and interpretation of the category system

The result of the process of inductive category-building is a system of categories on a specific topic which are connected to particular text modules. However, in the process of inductive category-building, it is not only necessary to classify certain text modules into the corresponding category, but it is also important to explain what is meant with the selected text module. The analysis and interpretation of the category system has the aim of making connections between the individual subcategories and can be conducted in three different ways (Mayring, 2015). Firstly, the whole category system can be interpreted with regard to the research question. Secondly, inductive main categories can be built with the help of a synthesizing content analysis. Deductive main categories can be built from insights gained from the theory. The approach of building deductive main categories was used in this thesis. Table 3 illustrates and summarizes the individual steps of building inductive categories in the process of a qualitative content analysis within this study.

3.4. Reporting and dissemination of the data

The last stage of a systematic literature review is the reporting and the dissemination of the results where practical recommendations are derived from the theoretical evidence (Tranfield et al., 2003). Tranfield et al. (2003) highlight that a good systematic review should enable practitioners to better understand the research by summarizing comprehensive primary research papers. In this thesis, the established category system was used to link the different themes with each other and to show contradictions as well as similarities. The justification and grounding of the conclusions was affiliated to the core contributions of the original material. Getting evidence into practice is the last point when writing a systematic literature review (Tranfield et al., 2003). This means that the insights and the conclusions which were gained from the systematic review need to be turned into guidelines for practice. The aim was therefore to create a reliable base of knowledge for practitioners by accumulating and synthesizing knowledge from a number of studies. The reporting and dissemination of the data can be found in the discussion of this thesis. In the discussion, the results of the systematic literature review will be connected to the literature in order to enlarge the existing knowledge base.

4. Results

This thesis conceptualizes ambidexterity as the simultaneous pursuit of exploration and exploitation. The analysis of the 40 most cited articles concerning the ambidexterity literature as well as six articles from practitioner literature reveal that there exist a variety of different solutions and mechanisms which can help to achieve ambidexterity in practice. An overarching element of every of these solutions is the role of the top management team (TMT). The results section is divided into three main categories, namely the measures at the TMT level needed to achieve ambidexterity, the actions which the TMT can take in order to achieve ambidexterity in their organizations, and the moderators and external factors affecting the pursuit of exploration and exploitation. Measures at the TMT level refer to those actions which directly concern the TMT and include factors such as the TMT constellation, leadership styles, strategic decisions and specific characteristics of leaders which are helpful for achieving ambidexterity. Approaches at the organizational level include different structural and contextual arrangements as well as human resource practices which should be implemented by managers in order to facilitate ambidexterity.

Lastly, the moderators and external factors are those elements which are more or less out of scope for decision-making and can only marginally be influenced by managers. These include environmental factors, the availability of resources, and certain characteristics of organizational networks which affect the strategies for pursuing ambidexterity. Other factors such as dynamic capabilities and the absorptive capacity of an organization will also be discussed in this regard. In order to ensure a differentiated exposition of the findings of the 40 most cited articles and the selected practitioner literature concerning the topic of ambidexterity, the results will be presented separately. This differentiated consideration should highlight possible differences concerning the solutions and mechanisms to achieve ambidexterity.

Table 3: Steps of a Qualitative Analysis of Content

(adopted from Mayring, 2015, pp. 62, 86).

1. Determination of the material
Analysis of how the material was produced
Characteristics of the material
Direction of the analysis
Theoretical differentiation of the research question
2. Procedure of the analysis
Determination of the procedure of the analysis (inductive category-building)
Working through the material
Selection of text modules pertaining to practical implications of ambidexterity
Determination and definition of categories
3. Coding and revision of the category system
Coding of the text modules
Subsumption or new building of categories
Revision of the categories after examining about 50% of the material
Final perusing of the material
4. Analysis and interpretation of the category system
Analysis with the help of the category system
Summary of the results
Interpretation into the direction of the research question

4.1. Implementing measures at the TMT level to achieve ambidexterity

As key leaders in an organization, senior executives play a major role in developing and strengthening ambidexterity (Raisch and Birkinshaw, 2008). Volberda et al. argue that "top management explicitly manages the balance of exploration and exploitation by bringing in new competencies to some units while utilizing well-developed competencies in others" (Volberda et al., 2001, p. 165). This section has the aim to provide a summary of the key mechanisms which are necessary to be implemented directly at the top management team level in order to achieve ambidexterity. First of all, these include the cognition of paradoxes which arise when balancing the conflicting and often contradictory agendas of exploration and exploitation, and the definition of a strategy which highlights the importance of ambidexterity. Secondly, different team constellations have an influence on how ambidexterity is managed at the TMT level. Therefore, different constellations of teams and their characteristics will be discussed. Thirdly, different characteristics of leaders and leadership styles which affect the achievement of ambidexterity will be analyzed. Lastly, it will be explained how formal structural and personal coordination mechanisms at the top management team level can affect ambidexterity.

4.1.1. Recognize and resolve paradox

The main task of managers and top management teams in an ambidextrous organization is to balance its short-term performance and its long-term adaptability (Smith and Tushman, 2005). This requires trade-offs regarding the allocation of resources as well as strategic decisions concerning negotiations between the existing and the new products in order to ensure the success of both agendas (Smith and Tushman, 2005). According to Smith and Tushman (2005) decisions can be made with regard to the distribution of resources between the existing product and the innovation and with regard to the recognition of opportunities and synergies arising from exploration and exploitation. "Exploration and exploitation require fundamentally different and inconsistent organizational architectures and competencies" (Smith and Tushman, 2005, p. 525). For the alignment and management of these different architectures, a top management team which can host these inconsistencies is needed (He and Wong, 2004). In order to balance the strategic decisions and to reduce or allow the coexistence of inconsistencies, the TMT needs to recognize and use this conflict (Smith and Tushman, 2005; Eisenhardt et al., 2010). Balanced decision making can thus be enabled by paradoxical cognition which is not only the acceptance of the presence of contradictory agendas, but also the embracing of contradictions and conflict (Smith and Tushman, 2005). The role of managers is therefore to support opposing forces and take advantage of the tensions between exploration and exploitation (Smith and Lewis, 2011).

Two main elements of coping with paradoxical tensions are acceptance and resolution (Smith and Lewis, 2011). The acceptance of paradox reduces the anxiety and stress which is associated with tensions and thus facilitates the implementation of resolution strategies (Smith and Lewis, 2011). The resolution of paradoxical tensions refers to the finding of so-

lutions on how to manage paradox. This can be done either through separating and choosing between the different tensions that occur or through seeking synergies that integrate contradictory agendas, such as exploration and exploitation, in order to foster sustainability (Smith and Lewis, 2011). To summarize the above, the management of paradox requires a senior leadership that tolerates and accepts the contradictions arising from multiple different agendas and that is able to resolve these tensions in order to increase the likelihood of ambidexterity (O'Reilly and Tushman, 2008). In other words, senior teams need to "recognize and translate different, ambiguous, and conflicting expectations into workable strategies" in order to "create integrative and synergetic value among exploratory and exploitative activities and to achieve organizational ambidexterity" (Jansen et al., 2008, p. 985). Eisenhardt et al. (2010) suggest three cognitive solution for balancing exploration and exploitation: abstraction, cognitive variety, and interruption. Abstraction implies that managers need to think abstractly in order to develop a common understanding of seemingly different opportunities. Cognitive variety can be defined as the multitude of different mental templates which help to generate a greater pool of solutions to accept and solve problems. Lastly, interruption facilitates exploration because a pause in the processes encourages rethinking, and also supports exploitation because a change in direction can assure to not waste time with non-working strategies.

Furthermore, the definition of distinct goals for exploration and exploitation can help to create paradoxical frames in order to enable positive conflict (Smith and Tushman, 2005). As a result, paradoxical frames do not only lead to reduced threat and anxiety, but they also enable teams to seize possible dualities and synergies between exploration and exploitation which, in turn, leads to greater performance (Smith and Tushman, 2005). Specifically, the management of contradictions which exist when pursuing both exploration and exploitation can be facilitated by two contrasting cognitive processes, namely differentiation and integration (Smith and Tushman, 2005). Differentiation at the top management level includes the clarification of differences regarding the strategy and organizational architectures, whereas integration involves the recognition of possible synergies between exploration and exploitation and between the strategy and organizational architectures (Smith and Tushman, 2005). Andriopoulos and Lewis (2009) similarly claim that a mix of both integration and differentiation tactics is necessary to cope with the paradoxes concerning exploration and exploitation. Firms need to manage these paradoxes at multiple levels which also requires that the different levels interact in order to enhance ambidextrous practices (Andriopoulos and Lewis, 2009).

Im and Rai (2008) have found that ontological commitment in an inter-organizational relationship positively influences explorative and exploitative knowledge sharing in the relationship. Ontological commitment refers to the ability of partnering firms to use digital boundary objects, meaning the usage of different information by differentiated units, to transfer, or to span, knowledge across boundaries (Im and Rai, 2008). Therefore, ontological commitment not only facilitates negotiation, but also help to create a common meaning of diverse interests, and thus to reconcile exploration and exploitation (Im and Rai, 2008). To be more precise, "the reliance on digital boundary objects should enable knowledge sharing by establishing standards for representation and transfer of data, facilitating interpretation of information, and promoting mutual discovery" (Im and Rai, 2008, p. 1285). In general, this means that managers need to identify potential benefits from contradictory forces and find synergies between them. The recognition, acceptance, and resolution of theses tensions allows managers to develop strategies for balancing exploration and exploitation and, thus, to achieve ambidexterity.

Findings from practitioner literature

In practitioner literature, the awareness of paradox is also considered as one of the most important aspects for balancing the two contradictory activities of exploration and exploitation and, as consequence, for achieving ambidexterity. Tushman et al. (2011) argue that senior managers need to embrace the tensions between new and old, the tension between the operating units and the core business at the top of the organization, which is enabled by a state of constant creative conflict at the top management level. Furthermore, managers need to embrace the inconsistencies which arise from hosting multiple, often conflicting, strategic agendas which are needed for exploration and exploitation (Tushman et al., 2011). This also involves the allocation and shifting of resources (e.g., financial resources and talent) between the innovation units and the core businesses, which requires the senior leaders to be "consistently inconsistent" (Tushman et al., 2011). Moreover, managers need to be able to reconfigure assets whenever there are changes in the competitive environment (O'Reilly and Tushman, 2011). This means that senior leaders not only need to sense these changes in technology, competition, and customer demands, but they also need to be able to respond to these changes in an efficient way (O'Reilly and Tushman, 2011). Additionally, senior teams have to enable both variety and local adaptation as well as collective action and strategic coherence (O'Reilly and Tushman, 2004). Tushman and O'Reilly (1996) highlight the importance of an alignment among strategy, structure, people, and culture. This means that the success of an organization is dependent on the leader's ability to increase alignment among strategy, structure, people, and culture during periods of incremental change which are disrupted by periods of discontinuous change which requires a simultaneous shift in the alignment of strategy, structure, people, and culture (Tushman and O'Reilly, 1996).

Conclusions

In general, both, findings from academic and practitioner literature, reveal that the cognition and management of paradoxes regarding exploration and exploitation are necessary for achieving ambidexterity in practice. Paradoxical cognition can, therefore, be regarded as the first step in imple-

menting ambidexterity in an organization. Certainly, it is the tasks of the top management team to be aware of the tensions that arise from balancing exploration and exploitation and to accordingly take appropriate actions to resolve these tensions. Therefore, managers need to evaluate how to achieve a balance between exploration and exploitation and make strategic decisions regarding the achievement of ambidexterity. The strategic elements of decisions regarding the implementation of ambidexterity in an organization will be explained in the next section.

4.1.2. Develop an ambidexterity-oriented strategy

One of the most important aspects for the coordination of exploration and exploitation is the implementation of an ambidexterity-oriented strategy. Such a strategy should be characterized by the presence of a shared vision among senior managers, as well as a common culture (Gibson and Birkinshaw, 2004; Jansen et al., 2008). Jansen et al. (2008) found empirical support that a senior team shared vision increases the achievement of organizational ambidexterity. Accordingly, a management's shared vision adds to the collective understanding of how senior team members might resolve the contrasting agendas of exploratory and exploitative activities and embodies common goals and shared values that provide for a common strategic intent which facilitates the reconciliation of contradictory agendas (Jansen et al., 2008). Additionally, shared values and common goals can help to overcome the problems arising from structural differentiation in ambidextrous organizations (Jansen et al., 2008). Simsek (2009) also notes that separate units responsible for exploration and exploitation are held together by a common strategic intent and an overarching set of values. This, in turn, enables the integration of exploration and exploitation at the top management team level (Andriopoulos and Lewis, 2009; Simsek, 2009).

O'Reilly and Tushman (2008) describe how sensing, seizing, and reconfiguring include important strategic decisions for the achievement of ambidexterity. First of all, sensing of opportunities involves scanning, searching, and exploration. This requires resources, routines and strategy-making processes that are linked to variation, resources that are used for competitive strategies and for sensing changes in the technological environment, as well as forums which give room for discussions of emerging chances (O'Reilly and Tushman, 2008). Apart from this, sensing also demands an open culture that promotes discussion, the commitment of financial and time resources by senior executives to enable long-term thinking, and a TMT that strengthens a long-term mindset and encourages exploration (O'Reilly and Tushman, 2008). Additionally, O'Reilly and Tushman argue that "to promote ambidexterity requires a senior management team that facilitates learning, challenges the status quo, accepts failure, and provides for the integration and transfer of knowledge, even as the exploitive subunit emphasizes the opposite" (2008, p. 190). Secondly, the seizing of opportunities is concerned with making appropriate strategic decisions and the execution of these strategies. Seizing thus demands "leaders who can craft a vision and strategy, ensure the proper organizational alignments (whether it is for exploitation or exploration), assemble complementary assets, and decide on resource allocation and timing" (2008, p. 191). This means that the TMT needs to have shared expectations concerning the strategic intent which should align the business model and the strategy (O'Reilly and Tushman, 2008). Thirdly, reconfiguring refers to the reallocation of resources from the mature businesses in the direction of new and emerging growth opportunities (O'Reilly and Tushman, 2008). Aside from structural decisions regarding the design of organizational systems and incentives in different units as well as the staffing of these separate units, senior leaders need to develop "processes by which these units are integrated in a value-enhancing way" (2008, p. 191). Reconfiguring therefore demands that leaders constantly realign their business along with the changes in the market in order to profit from ambidexterity (O'Reilly and Tushman, 2008).

In summary, the separate units for exploration and exploitation, which each consist of different competencies, systems, incentives, processes, and cultures, are internally aligned and held together by "a common strategic intent, an overarching set of values, and targeted structural linking mechanisms to leverage shared assets" (O'Reilly and Tushman, 2008, p. 193). There should be a clear consensus among the members of the TMT about the strategic intent, justifying the importance of ambidexterity (O'Reilly and Tushman, 2008). In order to enable ambidexterity in an organization, managers should communicate this strategy relentlessly and implement a common-fate incentive system (O'Reilly and Tushman, 2008). The articulation of a shared vision and a common set of values promote the establishment of a common identity which, in turn, fosters ambidexterity (O'Reilly and Tushman, 2008).

Additionally, when managers stimulate their firms to engage in corporate entrepreneurship, they need to have an entrepreneurial orientation (Dess and Lumpkin, 2005). Corporate entrepreneurship should have two major goals, namely the sensing and seizing of novel venture opportunities as well as strategic renewal. Entrepreneurial orientation can be defined as the strategy-making practices that firms utilize to sense and introduce corporate ventures. Dess and Lumpkin (2005) suggest that autonomy, innovativeness, proactiveness, competitive aggressiveness, and risk-taking are important determinants of a firm's entrepreneurial performance. Autonomy refers to independent actions by individuals or teams that are directed towards the vision of the organization. Innovativeness describes the willingness to experiment and to develop and introduce new products and services. Proactiveness can be defined as an aspirational perspective of an organization which is needed to anticipate future demands and to seize emerging opportunities. Competitive aggressiveness is aimed at the improvement of a firm's position in the marketplace and involves intense efforts to outperform other competitors. Lastly, risk-taking refers to the degree to which firms make decisions and take actions without being able to anticipate the possible consequences. These differ-

ent elements of entrepreneurial orientation can be integrated into the strategy for achieving ambidexterity and can act as guidelines for the whole organization to act ambidextrously.

Taylor and Helfat (2009) highlight the importance of organizational linkages to connect actors across different organizational units during period of technological transitions through communication and coordination. While coordination involves collaborative decision-making and planning for the allocation of resources, communication refers to the participation in discussions and meetings. The top management can enable middle managers to perform linking activities which have the aim to foster ambidexterity by "enabling firms to transition to a new technology while utilizing and adapting valuable preexisting capabilities that can be critical to the success of a transition" (Taylor and Helfat, 2009, p. 718). These managerial influences affecting the willingness of middle managers to carry out the organizational linking activities include economic, structural, social, and cognitive influences (Taylor and Helfat, 2009). An organization's cognition includes shared assumptions and understanding which can be found in the values, norms, and culture of an organization (Taylor and Helfat, 2009). In sum, this means that strategic decision-making involves the collaboration and interaction of the different members of the management teams in order to find solutions on how to balance exploration and exploitation in an organization.

Findings from practitioner literature

An overarching identity is key to achieving ambidexterity (Tushman et al., 2011). Tushman et al. (2011) argue that it is necessary for the TMT to develop strategic aspirations for the future and to create an emotionally compelling identity which is, at the same time, broad enough to give direction for the future. An overarching and compelling identity facilitates the simultaneous pursuit of opposing strategies, namely the exploitation of existing products and the exploration of new opportunities (Tushman et al., 2011). A common identity is supported by the articulation of a common vision and values across explorative and exploitative units (O'Reilly and Tushman, 2011). This, in turn, enhances trust, cooperation, and an aspirational point of view (O'Reilly and Tushman, 2011).

Furthermore, a compelling strategic intent, justifying the importance of both exploration and exploitation, facilitates the achievement of ambidexterity (O'Reilly and Tushman, 2011). A compelling strategic intent helps to avoid the success trap in that the short-term profits from exploration are not outweighed by long-term gains from exploitation activities (O'Reilly and Tushman, 2011). The senior team explicitly owns the strategy for achieving exploration and exploitation within the business units and have to relentlessly communicate this strategy across the entire organization in order to enhance cooperation and to avoid unproductive conflict (O'Reilly and Tushman, 2011). A common-fate reward system can help to implement a compelling strategy (O'Reilly and Tushman, 2011). Moreover, O'Reilly and Tushman (2004) claim that ambidextrous senior teams and leaders need to be committed to ambidexterity in order to be able to relentlessly communicate the vision. A clear and compelling vision provide direction towards the achievement of exploration and exploitation and highlights the need for ambidexterity and its benefits for all members of the organization (O'Reilly and Tushman, 2004).

In an ambidextrous firm, it is essential to create a culture that is simultaneously tight and loose (Tushman and O'Reilly, 1996). The tight aspect of a culture manifests itself in broadly shared norms which are needed for innovation, like for example autonomy, initiative, risk taking, and openness (Tushman and O'Reilly, 1996). The loose aspect of a corporate culture refers to the ways in which these shared norms are expressed and is dependent on the kind of innovation which is required (Tushman and O'Reilly, 1996). This means that organizations deploy multiple cultures, meaning subcultures in different business units, which are held together by a common overall culture (Tushman and O'Reilly, 1996). An overarching corporate culture facilitates the integration of the various subcultures, enhances information- and resource-sharing, enables consistency, and supports the creation of trust and predictability within a firm (Tushman and O'Reilly, 1996). Lastly, a tight-loose culture is encouraged by supportive leaders and a common vision who both support changes when they are needed (Tushman and O'Reilly, 1996).

From the above, it can be argued that a culture which is characterized by a shared set of cultural norms and values is vital for the management of paradox and, thus, for the alignment of exploration and exploitation. Tushman and O'Reilly (1996) exemplify this in the case of Apple where the employees who shared the same expectations about innovation, commitment, and speed, meaning who fit the values and who endorsed the cultural norms of the firm, stayed within the organization. However, although not all members of an organization need to act ambidextrously themselves, senior leaders that show resistance towards operating ambidextrously need to be dismissed (O'Reilly and Tushman, 2004). Additionally, Tushman and O'Reilly (1996) found that some firms, e.g., Nordstrom, derive their competitive advantage from a culture which is shared throughout the whole organization and which the competition can only imitate with difficulty. Furthermore, a social control system which allows the coordination of non-routine tasks during periods of change helps to establish a culture of shared norms and values (Tushman and O'Reilly, 1996).

O'Reilly et al. (2009) explain how ambidexterity relates to the establishment, the implementation, and the maintenance of new emerging business opportunities. Pertaining to this, the first phase is variation, meaning the establishment of a new idea or business opportunity. This includes the solicitation of a new idea from within or without the company, its socialization among senior executives and customers, and the assessment of the idea. The second phase is selection, meaning the implementation of the new idea. Frequent meetings of senior executives, a dedicated "A-Team Leadership", and the monitoring of progress are part of this phase. The last phase, retention, includes the moving from a future business to a growth business. The transition from a new idea into a

profitable business is heavily dependent on a strong leadership, a clear articulation of the strategy, early success in the market, as well as an aligned organization. O'Reilly et al. (2009) show that senior leaders need to continually reconfigure assets and sense the emergence of new business opportunities in order to be ambidextrous.

Conclusions

In sum, the development of an ambidexterity-oriented strategy includes the presence of a shared vision and a common culture. Again the results obtained from the academic and the practitioner literature are quite similar. While articles from academic literature primarily focus on the definition of shared values and goals which are embodied in an overarching vision, articles in practitioner literature highlight the importance of both, a compelling strategic intent and a common culture, which should facilitate the reconciliation of exploration and exploitation. It is again the responsibility of the top management team to collaboratively develop a strategy that is directed towards the achievement of ambidexterity and that highlights its importance. The strategic vision and culture need to be communicated to all members of an organization in order to establish common norms and values which enable the balance of exploration and exploitation and, as a consequence, the achievement of ambidexterity.

4.1.3. Ambidextrous leaders

There are certain characteristics which a leader needs to display in order to manage paradox and to, in turn, balance exploration and exploitation to achieve ambidexterity. Smith and Lewis (2011) state that two of these characteristics are cognitive and behavioral complexity as well as emotional equanimity. Raisch and Birkinshaw (2008) also suggest that leaders with complex behavioral repertoires are especially needed within contextual ambidexterity. These characteristics should help to accept paradoxical tensions and to take account of both/and possibilities (Smith and Lewis, 2011). O'Reilly and Tushman (2008) also argue that for the alignment of competencies, structures, and cultures to pursue either exploration or exploitation, a senior leadership team with the cognitive and behavioral flexibility to develop and nurture both is required. Effective leaders are therefore able to demonstrate complex behavioral repertoires that concurrently promote "consistency, stability, and control, as well as passion, courage, and wonder" (Gibson and Birkinshaw, 2004, p. 215).

Supportive and flexible leaders can be seen as key facilitators of organizational ambidexterity (Gibson and Birkinshaw, 2004). Jansen et al. (2008) state that in ambidextrous organizations, leadership behavior has an impact on the effectiveness of the senior team. This means that strategic executives may act more or less directive in integrating the paradoxical activities of exploration and exploitation (Jansen et al., 2008). Smith and Tushman (2005), for example, suggest that senior leader should designate different members of the senior team to pursue either explorative or exploitative activities, encourage them to be aware of the tensions, and to ease the discussion about possible synergies between the contrasting agendas.

Ambidextrous managers are also capable of fulfilling multiple roles at a time related to, for instance, the execution of both routine and nonroutine tasks (Adler et al., 1999), the carrying out of both creative and collective actions (Sheremata, 2000), or the conducting of tasks outside the narrow confines of their own job (Adler et al., 1999). In addition, ambidextrous managers are capable of both refining and renewing their knowledge, skills, and expertise (Sheremata, 2000). Hotho and Champion argue that "encouraging innovation requires a managerial mindset characterised by a positive, celebratory attitude towards innovation, combined with tolerance for failure, encouragement of open debate, and a prioritisation of innovation and change over stability and routine" (Hotho and Champion, 2011, p. 34). Therefore, a managerial mindset is based upon elements, such as flexibility, responsiveness to change, and room for creative thinking (Hotho and Champion, 2011).

Furthermore, managers can themselves display personal ambidexterity through the pursuit of both exploration and exploitation (Raisch et al., 2009). However, the level of personal ambidexterity is dependent on factors such as personal characteristics or organizational contexts in which a manager operates (Raisch et al., 2009). In general, Raisch et al. (2009) claim that the cumulative personal ambidexterity of an organization's members influences the ambidexterity of the whole organization, but that this is not the only determining factor.

Different leadership styles may also affect the achievement of ambidexterity. A transformational leadership style is, for instance, recommended by Hotho and Champion (2011) to enhance explorative innovation among employees. Jansen et al. (2008) also found that transformational leaders are necessary when it comes to encouraging critical debate and open discussion about conflicting demands among socially integrated teams. This means that transformational leaders are leaders that are respected and trusted, leaders with which followers can identify, leaders that are capable of motivating their followers to aspire to greater goals, and leaders that articulate a compelling vision (Jansen et al., 2008). Jansen et al. (2008) found empirical support that transformational leadership positively moderates the impact of senior team social integration on organizational ambidexterity, meaning that the relationship between senior team social integration and organizational is strengthened through the presence of transformational leadership. This is because socially integrated senior teams which are guided by a transformational leader are better able to reconcile tensions and to discuss conflicting task issues concerning exploration and exploitation (Jansen et al., 2008).

Findings from practitioner literature

The findings from the practitioner literature also reveal that leaders have a key function in achieving ambidexterity in an organization; or, as O'Reilly and Tushman state, "ambidextrous organizations need ambidextrous senior teams

and managers" (O'Reilly and Tushman, 2004, p. 81). Accordingly, there are some specific characteristics which leaders need to display in order to facilitate ambidexterity. Generally speaking, ambidextrous managers are sensitive to and understand the needs of their business (O'Reilly and Tushman, 2004). This requires them to "combine the attributes of rigorous cost cutters and free-thinking entrepreneurs while maintaining the objectivity required to make difficult tradeoffs" (O'Reilly and Tushman, 2004, p. 81). Tushman and O'Reilly (1996) describe ambidextrous leaders as those who revere the past but who, at the same time, have the willingness to continually change in order to meet future demands. Furthermore, the senior team has the ability to ensure that the entire organization has the willingness to learn from its competitive environment through the reinforcement of core organizational values such as autonomy, teamwork, responsibility, and innovation (Tushman and O'Reilly, 1996). Lower level managers often act restrained but also have to embody the culture in order to provide solutions directed at the interest of the organization (Tushman and O'Reilly, 1996). Although leaders have great autonomy in their actions, they are expected to deliver high performance in order to not be replaced (Tushman and O'Reilly, 1996).

Conclusions

From the above, it can be argued that certain characteristics of leaders influence the achievement of ambidexterity in an organization. Both, findings from practitioner and academic literature, suggest that leaders who are cognitively and behaviorally flexible and able to respond to the conflicting demands of exploration and exploitation are better able to implement ambidexterity. Therefore, a leader who is ambidextrous him or herself is likely to positively affect the achievement of ambidexterity in an organization because he or she is more capable of supporting other members of the senior team as well as other employees in the organization to act ambidextrously. Findings from the academic literature reveal that, in addition to this, a transformational leadership style might positively influence the achievement of ambidexterity because transformational leaders can motivate their followers to reconcile and discuss the tensions concerning exploration and exploitation.

4.1.4. Team constellations

The top management team has an important function in aligning contradictory agendas, such as exploration and exploitation. There are several elements which influence a TMT's effectiveness in achieving ambidexterity. Specifically, these concern not only certain characteristics of team, such as if a team is behaviorally integrated or if its members share common or diverse prior company affiliations, but also different constellations of a team which should facilitate the achievement of ambidexterity. This section will therefore provide a summary of the main elements which influence the composition of the senior team to be more efficient in the development of ambidexterity in an organization.

Behavioral integration

Lubatkin et al. (2006) regard a behaviorally integrated top management team as one of the major prerequisites for achieving ambidexterity in an organization. They found empirical evidence that behavioral integration positively influences both exploration and exploitation. Behavioral integration refers to a TMT construct that includes the level of a senior team's wholeness and unity of effort (Lubatkin et al., 2006). Behavioral integration depends on the level of the team's collaborative behavior, the quantity and quality of information exchanged, and the emphasis on joint decision-making. Therefore, if the top management team is behaviorally integrated, it gives senior managers the possibility to resolve conflicts, to communicate openly, and to develop a set of shared expectations and values which consequently facilitate ambidexterity (Lubatkin et al., 2006). Behavioral integration also leads to grater information sharing between the members of a team which is one critical factor needed for exploration and the discovery of new opportunities (Lubatkin et al., 2006). However, what might be necessary to enhance the behavioral integration among TMT members is a CEO who has the ability to select, motivate, evaluate, and coach the members of the TMT (Lubatkin et al., 2006).

Additionally, the level of a TMT's behavioral integration has a direct influence on how the team members manage contradictory knowledge processes related to exploration and exploitation, in that a greater integration increases the likelihood of simultaneously pursuing both activities (Lubatkin et al., 2006). While exploration involves a bottom-up learning process which requires that senior executives let go of old routines and move towards new opportunities, exploitation involves a top-down learning process which requires managers to institutionalize the routines that help to improve existing competencies (Lubatkin et al., 2006).

Pertaining to this, Jansen et al. (2009) do not refer to behavioral integration of the top management team, but to senior team social integration. Social integration refers to a team member's satisfaction in and attraction to the group as well as to the social interaction in a team which should increase collaborative problem solving and negotiation (Jansen et al., 2009; Jansen et al., 2008). While Jansen et al. (2008) could not find empirical support for that senior team social integration increases the achievement of organizational ambidexterity, Jansen et al. (2009) found empirical support for their hypothesis that senior team social integration mediates the relationship between structural differentiation and ambidexterity. They claim that "senior social team integration contributes to the mobilization and integration of operational capabilities at differentiated units to arrive at new combination of exploratory and exploitative activities" (Jansen et al., 2009, p. 801).

Common and diverse prior company affiliations

There are a variety of team constellations that help to manage the paradoxical tensions regarding exploration and exploitation, and thus to achieve ambidexterity. Beckman (2006) empirically studied the effect of the founding team composition on ambidexterity and found that the found-

ing team composition, and especially the team member's prior company affiliations, are important antecedents of exploratory and exploitative behaviors. More specifically, founding teams with common prior company affiliations rather display exploitative behaviors, while founding teams with diverse prior company affiliations rather display exploratory behaviors (Beckman, 2006). While common prior company affiliations include joint work experiences that foster a shared language and culture, as well as trust and mutual understanding among team members, diverse prior company affiliations offer more diverse networks and broader knowledge which are beneficial for innovation (Beckman, 2006). Teams who have both common and diverse prior company affiliations should have a common understanding to transfer knowledge and exclusive points of view to facilitate change and experimentation, thus displaying greater performance (Beckman, 2006). Therefore, a team whose members have both common and diverse prior company affiliations are more likely to simultaneously engage in exploration and exploitation and to, in turn, foster ambidexterity.

Leadercentric and teamcentric teams

Leadercentric and teamcentric teams are two possible team constellations that should facilitate the integration of strategic contradictions (Smith and Tushman, 2005). In leadercentric teams, it is the responsibility of the leader of a top management team to integrate exploration and exploitation (Smith and Tushman, 2005). Specifically, in leadercentric teams, it is the task of the leader to resolve paradoxical tensions which exist in top management teams whose members are assigned to different roles and goals, depending on whether they pursue explorative or exploitative activities. The leader is usually backed by supportive integrators, meaning one or more members of the TMT who bring in their expertise and skills and therefore help the leader to make balanced decisions. Moreover, extensive leader-member interactions should facilitate the exchange of knowledge between the leaders and the members of the TMT on order to better resolve the tensions existing between explorative and exploitative units. It is the leader's task in leadercentric teams to coach the team members to reinforce the differentiation between the strategic agendas of exploration and exploitation.

In teamcentric teams, a group of senior executives collaboratively integrate the contradictory agendas (Smith and Tushman, 2005). Because the members of teamcentric teams are responsible for different tasks concerning exploration and exploitation, it is necessary that the team is designed as a real team in which the members create paradoxical frames which provide them with shared mental models to develop a collective understanding of the paradoxical tensions and to enable them to clearly allocate the interdependent tasks. Furthermore, teamcentric teams should assign different roles, goals, and rewards at the product level as well as at the organizational level. While responsibilities at the product level should enhance the development of distinct and specific roles and information for either exploration or exploitation, responsibilities at the organizational level motivate members of teamcentric teams to consider the overarching strategic agendas necessary for integration. Teamcentric teams also benefit from frequent and high-quality team interactions in which team members benefit from the knowledge exchange with others. Teamcentric leaders have the responsibility to coach their members so that they actively handle conflict by focusing on both, their tasks at the product level as well as overarching issues at the organizational level. Lastly, Smith and Tushman (2005) suggest that a more democratic leadership style might be beneficial to teamcentric team, while a more authoritarian leadership style is more appropriate within leadercentric teams. In any case, Jansen et al. (2008) argue that strategic leadership can foster a senior team's effectiveness in ambidextrous organizations when senior leaders are promoted to work as a team.

Findings from practitioner literature

The findings from the practitioner literature also reveal that the top management team is responsible for making important strategic decision regarding the achievement of ambidexterity. Tushman et al. (2011) introduce two different approaches of how a team of leaders can be organized in order to enable the alignment of the innovation units and the core business at the top of the organization, namely hub and spoke teams and ring-team models. A hub and spoke team is characterized by a CEO who is placed at the center of the wheel. The business unit leaders surround this wheel heavily rely on the CEO who manages each spoke separately and communicate only with him, not with other business unit leaders. An inner circle consisting of two to three individuals servers as a point of interaction with the business unit leaders. The resolution of exploratory and exploitative strategies is the task of the CEO. In a ring-team model, on the other hand, decisions are made collectively together with the business unit leaders and the CEO. A ring-team model requires high communication, transparency, and collaboration in order to make decisions on how to allocate resources and make trade-offs between the present and the future. Both solutions enable the alignment of explorative and exploitative units at the top management level where the strategic decision are made.

Conclusions

This section shows that the constellation of a top management team affects the reconciliation of exploration and exploitation and that certain compositions of a TMT can lead to the facilitation of the achievement of ambidexterity. The findings from the academic literature show that certain characteristics of team, such as the behavioral integration of a TMT or the presence of common or diverse prior company affiliations, positively influence ambidextrous behavior. Regarding the more formal constellations of a team, which concern primarily its structure, both academic and practitioner literature provide interesting insights. While the academic literature suggests the use of teamcentric or leadercentric team constellations, the practitioner literature calls for hub and spoke or ring-team models as possibilities to facilitate ambidexterity.

Both approaches exhibit strong similarities with each other. Teamcentric teams can be compared with ring-team model constellations, as all team members collaboratively work together and display high levels of interaction and knowledge exchange. Leadercentric teams, on the other hand, can be compared with hub and spoke teams. Here, the leader is responsible for resolving paradoxical tensions and team members communicate and collaborate little with each other, but much with the leader. Apparently, both, teamcentric teams (or ring-team models) and leadercentric (or hub and spoke models), seem to be effective solutions in managing contradictory tasks, such as the alignment of exploration and exploitation. Consequently, the constellation which should preferably be implemented in an organization is likely to be dependent on different factors, such as the characteristics of the leader and his or her and leadership style, the organization's culture and vision, or other structural factors which may affect the constellation of the top management team.

4.1.5. Formal structural mechanisms and personal coordination mechanisms

There are certain formal and informal (personal) mechanisms which leaders can use to facilitate the achievement of ambidexterity in an organization or unit. While formal structural mechanisms include the level of a manager's decision-making authority, the centralization of decision-making, or the formalization of tasks within a business unit, personal coordination mechanisms refer to a manager's participation in cross-functional interfaces, or his or her connectedness to other members of the organization. How these mechanisms can be implemented in order to facilitate ambidexterity in an organization will be explained in this section. Also, it will be shown how different formal and informal mechanisms interact with each other to influence ambidexterity.

Formal structural mechanisms

One formal structural mechanism, namely the level of decision-making authority that a manager, affects the achievement of ambidexterity (Raisch et al., 2009). Raisch et al. (2009) claim that when an organizational context provides senior executives with decision-making authority, then this might result in higher levels of sense-making and cognitive processes that help to integrate exploratory and exploitative activities. Mom et al. (2009) have also found that a manager's decision-making authority positively relates to ambidexterity, because increased decision-making authority enhances a manager's self-control and ownership of tasks and decisions which, in turn, allows the manager to respond to different opportunities and to pursue a variety of different objectives. Jansen et al. (2009) also found empirical support that the higher a unit's centralization of decision making, the lower its level of exploratory innovation. Centralization of decision making refers to the level of authority and the concentration of decision making within a particular location of an organization (Jansen et al., 2009). However, centralization hinders individuals to sense and seize new and emerging opportunities which is detrimental to the requirements of exploration, including the solving of nonroutine problems (Jansen et al., 2009). The hypothesis that a unit's centralization in decision making positively influences exploitation could not be supported by Jansen et al. (2009).

Another formal structural mechanism is formalization (see Mom et al., 2009; Jansen et al., 2009). Jansen et al. (2009) found empirical support that the higher a unit's formalization, the higher its level of exploitative behavior. Formalization refers to the extent to which rules, procedures, and communications are formalized or documented (Khandwalla, 1977; as cited in Jansen et al., 2009). Formalization facilitates the codification of best practices which, in turn, leads to a more efficient exploitation and application of these practices (Jansen et al., 2009). Sheremata (2000) suggests that decentralized problem solving, reach in problem solving (i.e., the radius in which new ideas and information are searched inside or outside of organizational boundaries), and free flow of information are centrifugal forces that should facilitate exploration because they increase the likelihood that solutions will be found as well as the quality of solutions.

Personal coordination mechanisms

A manager's participation in cross-functional interfaces can be regarded as a personal coordination mechanism to achieve ambidexterity (Mom et al., 2009). Mom et al. (2009) found that the participation of a manager in cross-functional interfaces and his or her connectedness to other members of an organization is positively related to ambidexterity. Adler et al. (1999) state that cross-functional interfaces increase the trust between managers and enhances the ability to resolve conflicts concerning different needs, interests, and objectives of separate differentiated units. In addition to this, the participation in cross-functional interfaces allows managers to exchange knowledge which, as a consequences, enables them to renew and refine their existing knowledge base (Mom et al., 2009). Jansen et al. (2009) also found that cross-functional interfaces mediate the relationship between structural differentiation and ambidexterity. Cross-functional interfaces include platforms that facilitate knowledge exchange across explorative and exploitative units and help to build understanding and cooperation (Jansen et al., 2009).

Connectedness of a manager can be described as the extent to which a manager deploys networks of direct personal contacts to other members of other hierarchical levels or business units (Mom et al., 2009). Again, a manager with a large network of direct contacts has the possibility to acquire new knowledge through the exchange with other organizational members and can increase the trust and cooperation within the network (Mom et al., 2009). Dense social relations can, therefore, strengthen the collaborative resolution of tensions (Jansen et al., 2009). Jansen et al. (2006) empirically found support that the higher a unit's connectedness among its members, the higher its level of exploitative innovation. Connectedness helps organizational members to share and combine knowledge and to improve existing

knowledge bases which, as a consequence, is beneficial for exploitation (Jansen et al., 2006). Sheremata (2000) claims that connectedness, project manager influence, and cross-functional team influence are centripetal forces which should facilitate exploitation, because they speed up the problem-solving process and increase the quality of tradeoff decisions.

Combining formal and personal coordination mechanisms

Mom et al. (2009) also found positive interaction effects between formal structural and personal coordination mechanisms. For instance, they found that there is a positive interaction effect between a manager's decision-making authority and participation in cross-functional interfaces by the manager, on this manager's ambidexterity. This means that when a manager has decision-making authority over how and which tasks he or she performs and when this is accompanied by the cooperation with other managers of different units and functions, then this will have a positive effect on this manager's ambidexterity. The same is true for the interaction effect between a manager's decision-making authority and the connectedness of the manager to other organization members. In other words, when a manager increases his network, meaning his connectedness to others, this help him or her to better sense different opportunities which, in combination with increased decision-making authority, leads to increased ambidexterity (Mom et al., 2009).

In general, Mom et al. (2009) found that formal structural mechanisms are especially important for a manager's ambidexterity when this manager functions at an operational level. Cross-functional interfaces, on the other hand, are more conducive to a manager's ambidexterity when this manager functions on a business unit level, rather than on an operational level (Mom et al., 2009). Jansen et al. (2009) state that at the corporate level, social integration should be supported among the members of the top management team in an ambidextrous organization. At lower hierarchical level, more formal cross-functional interfaces are a means to foster the exchange of knowledge across explorative and exploitative units (Jansen et al., 2009). Lastly, a formal hierarchical structure, together with horizontal relationships, foster a manager's ambidexterity (Mom et al., 2009).

Conclusions

The results show that managers can use formal and informal coordination mechanisms to facilitate ambidexterity. The two main formal structural mechanisms include a manager's level of decision-making authority and the formalization of tasks in a business. Both, a manger's level of decision-making authority and the formalization of tasks, positively influence ambidexterity. Informal, or personal, coordination mechanisms include a manager's participation in cross-functional interfaces and his or her connectedness to other members of the organization. Both mechanisms do not only facilitate the exchange of knowledge between organizational members, but also increase the trust between a manager and others. This, in turn, leads to the ability to better resolve paradoxical conflicts and to achieve ambidexterity. Moreover, the findings show that a manager's decision-making authority positively interacts with the participation in cross-functional interfaces and the connectedness to other members of the organization. For managers, this means that when they succeed in having authority over decisions and when they deploy dense networks of direct personal contacts to other organizational members, they are more capable of acting ambidextrous themselves in order to implement measures to direct their organization towards ambidexterity.

4.2. Implementing organizational design solutions to achieve ambidexterity

The above section highlights that the role of leadership is to be aware of the paradoxical tensions existing between exploration and exploitation and to use the possibly synergies between them in order to achieve organizational ambidexterity (see Smith and Lewis, 2011). Furthermore, it was explained what leaders need to do by defining goals and strategies that help them make decisions. The question that arises from here is how leaders should operate within the organization to achieve ambidexterity? This section provides possible mechanisms and organizational design solutions that leaders can implement to support either certain structural or contextual arrangements, or specific human resource practices that should facilitate the implementation and development of ambidexterity.

4.2.1. Structural arrangements

As already mentioned earlier, structural (or partitional) ambidexterity has its roots in Duncan's work of 1967 who considers dual structures, one to initiate and one to execute, as viable options to ensure the long-term success of an organization (Simsek et al., 2009). According to Simsek (2009) and Jansen et al. (2009), the structural independence of the different units assures that exploitative and explorative activities do not 'stand in each other's way'. For example, through structural separation, the exploitative culture does not overwhelm the processes, structures, and cultures of the explorative units and the initiatives pursued in explorative units do not disrupt the exploitative activities in established units (Simsek, 2009). He and Wong (2004) argue that senior managers need to manage exploration and exploitation on a continuous basis, meaning simultaneously in a "steady-state perspective", for instance, through the implementation of a semi-structures design. Adler et al. (1999) refer to the creation of specialized units in order to engage in either routine or nonroutine tasks simultaneously as 'task partitioning'.

Specific features of explorative and exploitative units

Structural ambidexterity necessitates the creation of structurally independent units, each having its own strategies, structures, cultures, management teams, control and incentive systems (Benner and Tushman, 2003). While these two different logics are tightly coupled and integrated at the business unit level, they remain loosely coupled across different business units (Benner and Tushman, 2003). In other words, "within subunits the tasks, culture, individuals, and

organizational arrangements are consistent, but across subunits tasks and cultures are inconsistent and loosely coupled" (Benner and Tushman, 2003, p. 247). Within differentiation, the separate organizational units which are responsible for exploration are smaller, more decentralized, more flexible, and with loose processes and those responsible for exploitation are larger, less decentralized, and with tight processes (Benner and Tushman, 2003). Similar to the assertions of Benner and Tushman (2003), O'Reilly and Tushman note that the alignment of competencies, systems, structure, and culture to execute an explorative or exploitative strategy largely differ from one another. While exploitation necessitates "a short-term time perspective, efficiency, discipline, incremental improvement and continuous innovation", exploration demands "a longer time perspective, more autonomy, flexibility and risk taking and less formal systems and control" (O'Reilly and Tushman, 2008, p. 190).

Raisch et al. (2009) argue that organizations can implement structural solutions at different levels of an organization. Pertaining to this, a manufacturing plant may become ambidextrous by creating two distinct teams, one responsible for exploration and the other for exploitation (Raisch et al., 2009; Adler et al., 1999). A business unit may become ambidextrous by creating two subdivisions focusing on either exploration or exploitation (Raisch et al., 2009; Benner and Tushman, 2003). Lastly, a single team may become ambidextrous by giving each individual a distinct role (Raisch et al., 2009; Jansen et al., 2008). Organizations may also use structures to promote linking activities across different units (Taylor and Helfat, 2009). According to Taylor and Helfat (2009) structures have the aim to link and coordinate the activities of structurally interdependent units and involve rules, procedures, control systems, and coordination units.

Differentiation – integration tactics

Within structural ambidexterity, each unit operates independently. However, these differentiated units are organizationally interdependent with regard to the achievement of ambidexterity (Simsek et al., 2009). The coordination of exploration and exploitation therefore necessitates the integration at the top management team level (Benner and Tushman, 2003; Simsek et al., 2009). Structural differentiation, meaning the creation of subunits, allows organizations to explore and exploit. The top management team hereby serves as a point of integration to align these two contrasting domains (Smith and Tushman, 2005). Therefore, the organizational architecture requires highly differentiated units as well as top management team integration (Benner and Tushman, 2003; He and Wong, 2004; Gibson and Birkinshaw, 2004; Smith and Tushman, 2005). In other words, the coupling of exploration and exploitation can be achieved by the differentiation and integration of different explorative and exploitative projects (Andriopoulos and Lewis, 2009).

As dual structures can lead to difficulties in reconciling the explorative and exploitative activities of the individual units, strategic integration, and as a consequence organizational ambidexterity, can be reached by common aspirations among the top management team (Simsek, 2009). The integration at the senior team level enables a balanced allocation of resources and creates a cross-fertilization across exploitative and explorative units (Jansen et al., 2008; Smith and Tushman, 2005). The senior executives play a major role in the integration processes, the remaining organizational members are more or less segregated from the contradictory challenges of achieving ambidexterity (Simsek, 2009). However, Raisch and Birkinshaw (2008) suggests to also use lower-level integration mechanisms in order to promote the knowledge flows across different units. Still, there is consensus among a majority of researchers (e.g., Raisch and Birkinshaw, 2008) that the strategic integration across units is best achieved through the coordination at the top management level and a strong and overarching corporate culture. Tiwana (2008) found empirical support that knowledge integration at the project level improves alliance ambidexterity in innovation-seeking project alliances. Additionally, while strong ties are needed to integrate knowledge, bridging ties are required to assimilate new and diverse knowledge (Tiwana, 2008). Ambidexterity can be facilitated when strong ties complement bridging ties (Tiwana, 2008). Regarding the differentiation-integration tension, Cao et al. (2009) found that ambidexterity can be strengthened by close interrelations between existing and new knowledge. Raisch et al. (2009) explain that a synergistic effect between the new and the existing knowledge can be attained by employing the existing resources more fully in order to obtain new capabilities and by allowing new knowledge to be more fully integrated into the existing resource base.

Balance by unbalancing

As already mentioned earlier, firms usually encounter a natural drift towards efficiency, i.e. exploitation (Eisenhardt et al., 2010). Therefore, He and Wong (2004) argue that there needs to be a "counterbalance" between exploration and exploitation in order to continually adapt to changes in the environment. Eisenhardt et al. (2010) suggest three different structural mechanisms that should help to avoid this drift and to balance exploration and exploitation through unbalancing to favor flexibility, i.e. exploration. First of all, "simple rule strategies" consisting of heuristics allow for a quicker solving of problems and facilitate improvisation, thus supporting exploration (Eisenhardt et al., 2010). Secondly, managers can use simplification cycling, meaning the continual addition of new experiences to the structure and the simultaneous removal of structure, to not increase the structure (Eisenhardt et al., 2011). Thirdly, "flexibility-injecting structures", such as temporary tasks, prototypes, or alliances should facilitate exploration (Eisenhardt et al., 2011). Another mechanism to disrupt current balances between exploration and exploitation is radical innovation (O'Connor and DeMartino, 2006). O'Connor and DeMartino (2006) suggest that radical innovation, through the development of entirely new business lines, can lead to the creation of new markets. Hereby, each division hosts a proper infrastructure for radical innovation which should facilitate the investment in high-

uncertainty as well as high-risk projects in order to outperform other firms (O'Connor and DeMartino, 2006).

Findings from practitioner literature

The findings from the practitioner literature also suggest that the creation of small, decentralized, and autonomous units is one structural solution to achieving ambidexterity (Tushman and O'Reilly, 1996). When units are small and autonomous, employees can take greater risks and feel responsible for their own results (Tushman and O'Reilly, 1996). Similarly, O'Reilly and Tushman (2011) argue in favor of separate but aligned organizational architectures (i.e. business models, structure, incentives, metrics, and cultures) for the units responsible for exploration and exploitation and an integration of these different units at the senior team level. Again, the senior team has the responsibility to resolve the conflicts and tensions that stem from the implementation of these separate alignments (O'Reilly and Tushman, 2011, 2004).

The structural separation of exploitative and explorative units, however, implies that these units require different approaches in various fields. O'Reilly and Tushman (2004) summarize the major differences. First of all, the main tasks in exploitative units concern operations, efficiency, and incremental innovation, whereas in explorative units they concern adaptability, new products, and discontinuous innovation. Secondly, regarding the required competencies, those in exploitative units are operational, while in explorative units they are entrepreneurial. Thirdly, there is a formal and mechanistic structure in exploitative units, whereas the structure in explorative units is adaptive and loose. Fourthly, in exploitative units, margins are controlled and productivity is rewarded, while in explorative units, milestones are controlled and growth is rewarded. Fifthly, while in exploitative units the culture emphasizes efficiency, low risk, quality, and customers, the culture in explorative units is directed towards risk taking, speed, flexibility, and experimentation. Lastly, concerning the role of leadership, in exploitative units it is more authoritative and top down, whereas in explorative units it is visionary and involved.

The findings from the practitioner literature, like those from the academic literature, suggest that organizations should 'balance by unbalancing'. In order to achieve a balance between exploration and exploitation, Tushman and O'Reilly (1996) argue that managers might need to "cannibalize" their own business. This means that there needs to be a disruption of the established alignment between exploration and exploitation in order to respond and adapt to changing environments. Therefore, old structures, processes, and systems which might have led to the short-term alignment of strategy, structure, and culture need to be destroyed and replaced with new ones when changes in the competitive environments require modifications within the organizational architectures.

Conclusions

It is widely acknowledged that the structural separation of differentiated business units operating in either exploration or exploitation is a feasible solution for achieving ambidexterity. The top management team has a vital function in integrating these structurally separate units in order to assure the reconciliation of exploration and exploitation. Both, findings from practitioner and from academic literature support the mechanisms of differentiation and integration for the achievement of ambidexterity. Furthermore, the results from articles of the academic and the practitioner literature both reveal that the different units for exploration and exploitation need to be managed by implementing substantially different processes, structures, tasks, competencies, and cultures. Certainly, it is the responsibility of the top management team to implement these diverse elements and to make sure that the contrasting units and elements remain aligned with each other.

4.2.2. Contextual arrangements

While structural solutions involves the creation of separate units to pursue exploration and exploitation, contextual solutions allow the simultaneous pursuit of exploration and exploitation within the same unit (see Simsek et al., 2009). This type of ambidexterity requires "the creation of a context that promotes a behavioural orientation towards a combined capacity for both exploitation and exploration, one in which they can 'simultaneously flourish'" (Gibson and Birkinshaw, 2004, p. 209). Therefore, when contextual ambidexterity has successfully been implemented in organization, individuals in a business unit can generate value for existing customers and simultaneously are able to sense and react to changes in the task environment (Gibson and Birkinshaw, 2004). Through a context which equally emphasizes high performance (a combination of discipline and stretch) and social support (a combination of support and trust), ambidexterity is facilitated through the ability of individuals "to make integrative judgements as to how to best divide their time between the conflicting demands for alignment and adaptability" (Gibson and Birkinshaw, 2004, p. 211). Performance management is concerned with encouraging employees to voluntarily aim at high and ambitious goals, while social support ensures that employees establish these goals in the context of a cooperative work environment which is characterized by support and trust by others (Gibson and Birkinshaw, 2004).

How the four behavior-framing attributes, namely discipline, stretch, support, and trust, can be developed and reinforced by managers (see Ghoshal & Bartlett, 1994 in Gibson and Birkinshaw, 2004) will be explained now. First of all, the attribute of discipline can be defined as the voluntary attempt of individuals to meet all the expectations which arise from their explicit or implicit obligations. Discipline can be achieved by the "establishment of clear standards of behavior and performance, a system of open, candid, and rapid feedback, and consistency in the application of sanctions" (p. 213). Secondly, stretch can be described as the endeavor of individuals to orient themselves towards more ambitious goals than less ambitious ones. Stretch can be created by the "establishment of a shared ambition, the development of a

collective identity, and the ability to give personal meaning to the way in which individuals contribute to the overall purpose of an organization" (p. 213). Thirdly, the attribute of support refers to the instigation of individuals to lend assistance and encouragement to other members. Support can be developed by "[m]echanisms that allow actors to access the resources available to other actors, freedom of initiative at lower levels, and senior functionaries giving priority to providing guidance and help rather than to exercising authority" (p. 213). Fourthly, and lastly, trust induces individuals to mutually rely on each member's commitments. Trust is established by "[f]airness and equity in a business unit's decision process, involvement of individuals in decisions and activities affecting them, and staffing positions with people who possess and are seen to possess required capabilities" (p. 213).

According to Gibson and Birkinshaw (2004), the establishment of a high performance behavioral context necessitates managerial guidance concerning transparency in accessing resources, autonomy to take initiatives, and equity and fairness in decision-making processes. Simsek (2009) states that the behavioral context of organizational members is defined by a carefully selected set of systems and processes which enables individuals to perform both exploitative and exploratory activities.

Im and Rai (2008) state that contextual elements of the overall management system include systems, processes, and beliefs that align specific behavior, like for instance knowledge sharing, in an inter-organizational relationship. They claim that "contextual ambidexterity is the nonsubstitutable combination (i.e., interaction) of alignment and adaptability of the management system that includes service level arrangements, incentives, and planning and review meetings that govern a relationship" (Im and Rai, 2008, p. 1284). Accordingly, in a long-term inter-organizational relationship, contextual ambidexterity should foster both exploitative and explorative knowledge sharing (Im and Rai, 2008). The authors found empirical evidence that the greater the contextual ambidexterity in an inter-organizational relationship, the greater the exploitative and the explorative knowledge sharing in the relationship.

Findings from practitioner literature

Pertaining to contextual ambidexterity as the implementation of an organizational context including two dimensions, namely performance management (a combination of stretch and discipline) and social support (a combination of support and trust) Gibson and Birkinshaw (2004) present a counterarticle in the MIT Sloan Management Review (see Birkinshaw and Gibson, 2004). As mentioned above, while performance management refers to the incitement of individuals to deliver high-quality results, social support aims at giving individuals security and room to accomplish their tasks. In their article in the practitioner review, Birkinshaw and Gibson (2004) provide more concrete guidelines for managers on how to build an ambidextrous organization with the help of context. First of all, managers need to diagnose the current position of their organizational context in terms of performance management and social support. Secondly, managers should just focus on a few levers, but employ those consistently. Thirdly, managers should build understanding through a clear and consistent communication of the message throughout the organization. Fourthly, contextual ambidexterity and structural ambidexterity do not exclude one another, rather they are complements. While structural separation may be a good starting point for new initiatives, the integration with the mainstream organization should be conducted as quickly as possible through contextual ambidexterity. Lastly, contextual ambidexterity should be viewed as "driving leadership", meaning that individuals make their own choices about how and where to focus their capacities through the creation of a supportive context (p. 55). In this way, leadership is not only displayed by the leaders themselves, but also by every other member of the organization. This assumption also leads to the prerequisite that an organization needs ambidextrous employees (see Birkinshaw and Gibson, 2004).

Birkinshaw and Gibson (2004) emphasize four attributes of ambidextrous individuals in an organization.Firstly, ambidextrous employees are proactive and can seize opportunities outside the confines of their job and consequently act in the broader interests of the organization. Secondly, they are cooperative and able to work with others to increase their efforts. Thirdly, ambidextrous employees have the capability to build internal linkages. Fourthly, and lastly, they are multitaskers able to do more than one task at a time. These characteristics allow employees not only to act spontaneously and independently, but also to quickly adapt to emerging opportunities and to take actions that are aligned with the overall strategy of the organization (Birkinshaw and Gibson, 2004).

Conclusions

Compared to structural ambidexterity, the creation of a context which is characterized by four-behavior framing attributes, namely discipline, stretch, support, and trust is an alternative solution for achieving ambidexterity in an organization. Within contextual ambidexterity, the focus lies on the individual level, meaning that employees have the possibility to individually switch between explorative or exploitative activities, given that the context provides the necessary requirements. The findings retrieved from the practitioner literature reveal that, when pursuing a contextual ambidexterity approach, it not only requires ambidextrous leaders, but also ambidextrous employees who display certain characteristics in order to act ambidextrously. Furthermore, while the results obtained from the academic literature specifically highlight how managers can develop and reinforce the individual elements of the context, i.e., discipline, stretch, support, and trust, the findings from the practitioner literature provide more general implications which can be regarded as necessary prerequisites for building an ambidextrous organization with the help of context.

4.2.3. Human resource practices

Human resource practices are one means to facilitate the achievement of ambidexterity in an organization. Therefore, this section will give an overview of the main human resource practices which can be implemented in an organization to achieve ambidexterity. More specifically, it will be explained how job enrichment, training, incentives and rewards, as well as certain human resource practices for the development of intellectual capital can be used to foster ambidexterity in an organization. Job enrichment and training have the aim to enable employees to perform both, exploration and exploitation, at the same time. Incentives and contingency rewards are especially used among senior teams to enhance their cooperation in aligning these contradictory agendas. Lastly, the development of intellectual capital should help to acquire and integrate knowledge within an organization.

Job enrichment, training, and socialization

There are a variety of human resource (HR) practices which facilitate the simultaneous pursuit of exploration and exploitation. In contextual ambidexterity, one of these practices includes job enrichment (Adler et al., 1999). With the help of job enrichment programs, for example, employees are provided with training and experience in exploration and exploitation respectively, which enables them to perform both sets of activities (Simsek et al., 2009) and to become more innovative in their routine tasks (Adler et al., 1999; Gibson and Birkinshaw, 2004). Other organizational factors, such as socialization, recognition, or team-building practices are also suggested to affect individual's ability to think and act ambidextrously (Ghoshal and Bartlett, 1997; as cited in Raisch et al., 2009). Taylor and Helfat (2009) claim that social rewards, such as emotional support, attachment to a group, self-esteem, or social status have an effect on how well individuals are able to establish linkages across interdependent organizational units. Adler et al. (1999) state that two important contextual factors are training and trust. Training ensures that individuals have sufficient knowledge, skills, and abilities to engage in mechanisms such as job enrichment or task partitioning (Adler et al., 1999). Trust regarding the consistency of task completion by others and the trust in goal congruence especially facilitates the implementation of meta-routines and other mechanisms, such as job enrichment (Adler et al., 1999).

Incentives

Rewards and other incentives are another means to facilitate ambidexterity in an organization. Jansen et al. (2008), for example, empirically found that senior teams receiving team contingency rewards were better able to pursue ambidexterity. Kaplan and Henderson (2005) claim that, especially during technological periods of change, managers need to make decisions about what to reward. In order to effectively implement a new technology for instance, monetary compensation and salary-increasing promotions are a means to promote organizational linkages (Kaplan and Henderson, 2005; Taylor and Helfat, 2009). However, due to the tight linkage between cognitive frames and incentives, a change in one must always be accompanied by a change in the other (Kaplan and Henderson, 2005). Jansen et al. (2009) suggest that senior team contingency rewards are motivational in that they help senior managers to participate in problem solving and decision making. Contingency rewards depend on the extent to which the team's outcomes have an effect on individual team members' benefits (Jansen et al., 2009; Jansen et al., 2008). Contingency rewards can be used to strengthen the cooperation among senior team members which are responsible for different explorative and exploitative activities, to create commitment to different goals, and to resolve tensions regarding the allocation of resources (Jansen et al., 2009; Jansen et al., 2008). Smith and Tushman (2005) claim that senior team contingency rewards can also drive leaders to surmount their direct interests and to reallocate resources in order to reach establish an integrative value across the units responsible for exploitation and exploration. In sum, Jansen et al. argue that "compensating senior team members for overall firm performance decreases the chance of interest asymmetries and encourages senior team members to seek opportunities for strategic synergies across inconsistent exploratory and exploitative organizational units" (Jansen et al., 2008, p. 999).

Hotho and Champion (2011) argue that extrinsic motivation incentives are rather counterproductive to the innovativeness and productivity of employees. Rather, managers should provide intrinsically motivating incentives, which allow employees to innovate and to learn new things (Hotho and Champion, 2011). Innovation and knowledge creation can furthermore be facilitated by "flexibility, networked flatter structures, self-organising teams and projects, devolved decision making and democratic lines of communication" (Hotho and Champion, 2011, p. 34). These measures should enable individuals to work autonomously and to create a sense of ownership for their work (Hotho and Champion, 2011). Moreover, feedback and rewards should be adjusted to the process of work, and not simply the results of the work (Hotho and Champion, 2011). The tolerance of failure and the encouragement of risk by managers should further engage employees in experimentation and innovation (Hotho and Champion, 2011).

Development and management of intellectual capital

Kang and Snell (2009) describe how different types of intellectual capital affect either exploration or exploitation and which human resource practices can be implemented to develop intellectual capital. As the authors provide very detailed human resource management practices, the main findings concerning intellectual capital (i.e. specialist and generalist human capital, cooperative and entrepreneurial social capital, and mechanistic and organic organizational capital), their effect on exploration and exploitation, and their implementation in practice will be described here.

To begin with, specialist human capital is more likely to focus on exploration because it includes domain-specific knowledge which can be used to acquire new knowledge

(Kang and Snell, 2009). Generalist human capital, on the other hand, is more likely to focus on exploitative learning because it deploys multiple knowledge domains (Kang and Snell, 2009). Cooperative social capital is likely to facilitate exploitation as it includes strong and dense network connections which enables the assimilation of knowledge (Kang and Snell, 2009). Entrepreneurial social capital facilitates explorative learning as a looser connected social system enhances a firm's flexibility which is needed to expand and acquire new knowledge (Kang and Snell, 2009). Mechanistic organizational capital should facilitate exploitation and includes standardized structures and processes and routines which enable coordination (Kang and Snell, 2009). Lastly, organic organizational capital should facilitate exploratory learning through simple routines, structures, and cultures which leave room for autonomy and experimentation (Kang and Snell, 2009).

Given the above assumptions, exploration is likely to be facilitated by an intellectual capital architecture including generalist human capital, entrepreneurial social capital, and organic organizational capital, whereas exploitation is likely to be supported by an intellectual capital architecture that includes specialist human capital, cooperative social capital, and mechanistic organizational capital (Kang and Snell, 2009). Aside from this, there also exist two hybrid forms which should support ambidextrous organizational learning: refined interpolation and disciplined extrapolation. Refined interpolation is a combination of specialist human capital, providing the expertise required for exploitative learning, cooperative social capital, which helps the specialists to share and integrate knowledge for deeper exploitation, and organic organizational capital, which promotes the continuous integration of diverse knowledge bases (Kang and Snell, 2009). Disciplined extrapolation combines generalist human capital and entrepreneurial social capital which facilitate exploration through the acquisition and sharing of knowledge and mechanistic organizational capital, which ensures that new knowledge bases can be integrated efficiently (Kang and Snell, 2009). The organizational capital hereby always has the purpose of transforming people-embodied knowledge (i.e. knowledge embodied in human or social capital) into organizational knowledge (Kang and Snell, 2009).

Having explained that different compositions of intellectual capital architectures can support ambidextrous organizational learning, the question arises how the different intellectual capital types can be managed with the help of human resource practices. Developing generalist human capital requires a 'skill-based development' including broad and multidimensional job designs, job rotations, recruiting based on potential, extensive training, with skill-or knowledge-based inventive systems (Kang and Snell, 2009, p. 79). For developing specialist human capital, a 'job- or function-based development' including narrow job designs, focused career development, recruitment based on the fit between persons and jobs, intensive training for the improvement of job-related skills, with incentive systems that focus on individuals' performance and effort in current jobs for compensation is suggested by Kang and Snell (2009, p. 80). The development of cooperative social capital can be managed by internal labor market (ILM)-based systems including internal staffing/promotion, seniority-based compensation (including fixed bonus and egalitarian pay structure), socialization (e.g. mentoring, person-organization fit criteria for recruiting and promotion, extensive orientation, team structures, or multi-source feedback) (Kang and Snell, 2009, p. 80). For developing entrepreneurial social capital, market- or network-based employee relations systems including extensive external staffing that utilizes various external sources of HR, performance-based compensation (e.g. individual incentives, pay for reputation, hierarchical pay structure), and general development experiences (e.g. cross-training, training for interpersonal skill improvement, social events) may be appropriate HR practices (Kang and Snell, 2009, p. 80). The development of mechanistic organizational capital can be managed through performance/control systems targeted towards "error avoidance", behavior-based evaluation and rewards, specific behavioral appraisal systems (e.g. behavioral observation scales), and performance program that are imposed top-down (Kang and Snell, 2009, p. 81). For developing organic organizational capital, performance/control systems targeted towards "error embracing", the reduction of status barriers between managers and employees, employees' participation in problem-solving and decision-making, extensive transference of tasks and responsibilities to employees, providing chances to use personal initiatives, encouraging and implementing employee suggestions, and developmental performance appraisal can be used (Kang and Snell, 2009, p. 81).

Conclusions

This section provides a summary of the main human resource practices which can be implemented in order to facilitate ambidexterity. Managers can, for instance, implement certain trainings or job designs, such as job enrichment or task partitioning, to enhance ambidexterity among employees. These solutions should enable employees to better perform exploration and exploitation simultaneously. Furthermore, incentives and rewards can help to foster ambidexterity within an organization. Contingency rewards are especially used within senior teams to enhance trust and cooperation for the exchange of knowledge. This should help senior teams to align and reconcile the contrasting demands of exploration and exploitation. This section also gave an overview of the human resource practices which are important for the development of intellectual capital in an organization. Intellectual capital is an important intangible asset of an organization and can be divided into human, social, and organizational capital (Kang and Snell, 2009). 'Skill-based development' and 'job-/function-based development' can be used to develop generalist and specialist human capital, respectively. 'ILM-based systems' and 'market- or network-based employee relations systems' should facilitate the development of cooperative and entrepreneurial social capital. Lastly, performance/controls systems targeted

either towards 'error avoidance' or 'error embracing' can be used for the development of either mechanistic or organic organizational capital.

4.3. Moderators and external factors influencing the achievement of ambidexterity

This last section will provide a summary of moderators and external factors which influence the achievement and development of ambidexterity. These factors are more or less out of scope for decision-making and can often only be marginally controlled by managers. First of all, it will be explained how the availability of organizational resources affects the management of ambidexterity. Secondly, a variety of environmental factors will be introduced and suggestions of how managers can respond to these will be provided. Thirdly, the influence of different characteristics of an organization's network will be evaluated with regard to their effect on ambidexterity. Fourthly, it will be described how externalization and absorptive capacity can influence the management of ambidexterity. Lastly, it will be explained how dynamic capabilities and routines can help an organization to facilitate ambidexterity.

4.3.1. Manage ambidexterity in consideration of the availability of resources

Several authors have investigated the effect of resource endowment on ambidexterity (e.g., Voss et al., 2008; Cao et al., 2009). Voss et al. (2008) for example found empirical evidence for a positive relationship between organizational slack as well as human resource slack and product exploitation. Furthermore, they found that when an environment is perceived as more threatening, the relationship between financial slack and product exploration becomes more negative and the relationship between customer relational slack and product exploitation becomes less negative.

In general, structural ambidexterity, meaning the decoupling of organizational units (Benner and Tushman, 2003), should be pursued in larger firms with high levels of unabsorbed slack which operate in threatening environments (Voss et al., 2008). Similarly, Lubatkin et al. argue that small firms "lack the amount of slack resources and the kind of hierarchical administration systems that can help or impede larder firms in managing their contradictory knowledge processes and, thus, affect the attainment of ambidexterity" (Lubatkin et al., 2006, p. 647). Therefore, structural ambidexterity may be the more appropriate solution for large and diversified firms, whereas leadership-based ambidexterity may be more beneficial to smaller or more focused firms. Lubatkin et al. (2006) explain this by stating that there are fewer hierarchical levels in smaller firms which enables managers to engage in strategic and operational roles which require both exploration and exploitation.

Contextual ambidexterity, meaning "the behavioral capacity to simultaneously demonstrate alignment and adaptability across an entire business unit" (Gibson and Birkinshaw, 2004, p. 209), is the more appropriate solution for smaller firms or for single business units in a large firm who may not have the resources available to deploy multiple subunits and who operate in more stable environments (Voss et al., 2008). Raisch and Birkinshaw (2008) suggest that rich firms usually have the resources to both explore and exploit simultaneously, while firms with less resources may not be capable of implementing such a strategy and may therefore focus on a one-sided orientation on either exploration or exploitation. This argument is supported by Lin et al. (2007) who have found empirical evidence that large firms will tend to benefit more from an ambidextrous formation of exploratory and exploitative alliances, whereas a smaller firms will tend to benefit more from a focused formation of either exploratory or exploitative alliances. This can be explained by the fact that larger firms are exposed to relatively loose resource constraints which enables them to allocate large quantities of their resources to either explorative or exploitative activities (Lin et al., 2007).

Cao et al. (2009) found that whether a firm pursues a balance of a combined dimension should depend on the availability of resources. The balance dimension seeks to achieve ambidexterity through a relative balance between exploration and exploitation and is more beneficial to resource-constrained firms (Cao et al., 2009). In other words, small firms with little resources benefit from a trade-off between exploration and exploitation. The combined dimension seeks to a achieve ambidexterity through the combined magnitude of exploration and exploitation and is more beneficial to firms with greater access to resources (Cao et al., 2009). In other words, large firms which operate in munificent environments and have access to internal and/or external resources benefit from combining high levels of exploration and exploitation simultaneously.

Conclusions

The results from the academic literature show that the availability of resources in an organization has an effect on which strategy should be pursued to best achieve ambidexterity in an organization. Pertaining to this, the structural separation of explorative and exploitative business units is the better solution for larger firms who possess high levels of unabsorbed slack. Contextual ambidexterity, on the other hand, may be the better strategy for smaller firms who might not have sufficient resources to enable the pursuit of exploration and exploitation in different subunits. Therefore, smaller firms should concentrate on implementing a strategy that focuses on the development of employees to simultaneously pursue exploration and exploitation within a business unit. In sum, this means that managers need to adjust their strategy for achieving ambidexterity towards the endowment of organizational resources that their firm possesses.

4.3.2. Manage ambidexterity in consideration of different environmental factors

The nature of the environment in which an organization operates may sometimes affect the choices made in strategic decisions regarding ambidexterity. Uotila et al. (2009)

for example found that the research and development intensity of the industry in which firms operate positively moderates the relationship between the relative amount of exploration orientation and the financial performance of the firm. Moreover, different levels of technological dynamism require a different balance of exploration and exploitation in order to enhance performance (Uotila et al., 2009). This implies that firms facing an environment with lower technological dynamism might concentrate their efforts towards exploitation, whereas high technological dynamism requires firms to pursue sufficient exploration in order to avoid inertia through an overemphasis on exploitation (Uotila et al., 2009).

Sidhu et al. (2007) explored the moderating effect of environmental dynamism on the relationship between nonlocal supply-side search, demand-side search, and spatial search and innovativeness, meaning the ability to successfully introduce new products and services. The authors found that environmental dynamism moderates the relationship between innovativeness and the amount of nonlocal supply-side search and demand-side, so that when there is low dynamism this relationship is positive and when there is high dynamism, the relationship is negative. Furthermore, environmental dynamism does not influence the positive relationship between greater spatial search and innovativeness. However, not only the amount of search matters, but also its context (Sidhu et al., 2007). More precisely, this implies that when firms operate in a highly dynamic context they should pursue supply-side and spatial-search exploration together with demand-side exploitation (Sidhu et al., 2007). When the environment is stable, firms benefit from combining demand-side and spatial search exploration with supply-side exploitation (Sidhu et al., 2007). In general, the greater the amounts of nonlocal supply-side, demand-side, and spatial search, the greater the innovativeness of the firm (Sidhu et al., 2007).

Additionally, Simsek (2009) suggests that environmental dynamism as well as environmental complexity both positively moderate the relationship between organizational ambidexterity and organizational performance. Environmental dynamism is characterized by the rate of change as well as the extent of instability in an environment (Dess & Beard, 1984; as cited in Jansen et al., 2009). In dynamic environments, acting ambidextrously enhances an organization's ability to quickly react to actions of competitors and to customer demands, while in stable environments it might be more beneficial to focus on either exploration or exploitation to sustain a competitive advantage (Simsek, 2009). In complex environments, organizational ambidexterity can be used to handle both explorative and exploitative activities in order to enhance performance (Simsek, 2009). Jansen et al. (2006) found empirical support that environmental dynamism positively moderates the relationship between exploratory innovation and financial performance, and that environmental dynamism negatively moderates the relationship between exploitative innovation and financial performance. Dynamism rapidly makes existing products and services obsolete, explaining the negative relationship with exploitation (Jansen et al., 2006). Furthermore, Jansen et al. (2006) found that environmental competitiveness positively moderates the relationship between exploitative innovation and financial performance. Environmental competitiveness refers to the degree of competition in an external environment (Jansen et al., 2006). When a business is able to expand its current and existing products and services in a competitive environment, it can increase its customers' loyalty and hence defend its market position in order to increase its financial performance through exploitative innovation (Jansen et al., 2006).

In sum, Jansen et al. (2006) empirically found that in dynamic environments, it is more beneficial to pursue exploratory innovation, whereas in competitive environments, it is more beneficial to pursue exploitative innovation in order to enhance a unit's financial performance. Lin et al. (2007) have also found empirical support for their hypothesis that a firm with an ambidextrous formation of exploratory and exploitative alliances will tend to exhibit better performance in an uncertain environment, whereas a firm with a focused approach on either exploration or exploitation will tend to have better performance in a stable environment. An ambidextrous design therefore is able to strategically balance the different demands of exploration and exploitation which leads to enhanced performance in uncertain environments (Lin et al., 2007).

Eisenhardt et al. (2010) claim that leaders who are effective in balancing exploration and exploitation should avoid highly ambiguous environments and rather structure uncertain environments to their advantage. In highly unpredictable environments, managers should keep the structure minimal and undertake flexible adjustments of the structure as the situation requires (Eisenhardt et al., 2010). Furthermore, managers often face multiple environments. On the one hand, they have to manage for efficiency and exploitation in the existing market, and on the other, managers need to foster flexibility and exploration in their new and dynamic market (Eisenhardt et al., 2010).

Conclusions

The results obtained from the academic literature show that certain characteristics of the environment influence the strategies for pursuing either exploration, exploration, or both. According to this, ambidexterity, meaning the simultaneous pursuit of exploration and exploitation, may positively influence a firm's performance when this firm operates in dynamic or complex environments. Contrary to this, in stable environments firms can also consider to implement a one-sided focus on either exploration or exploitation. In sum, managers need to be aware of changes in their environment and take actions accordingly. So, on the one hand, managers need to defend their competitive position in existing markets through exploitation and, on the other hand, they need to establish their position in emerging and dynamic markets through exploration.

4.3.3. Manage ambidexterity in consideration of different network characteristics

Networks, meaning the connections which organizations have to others, may affect organizational ambidexterity. For instance, Simsek (2009) examines how network centrality and diversity affect ambidexterity and how this relationship is influenced by a variety of other moderators. He proposes that network centrality, meaning the extent to which an organization is well connected to others in a network, has a curvilinear relationship with organizational ambidexterity. This means that a moderate level of network centrality, having neither too low nor too high levels of centrality, should have the most positive effects regarding ambidexterity (Simsek, 2009). Simsek (2009) furthermore suggests that this relationship can be strengthened when an organization deploys dual structures (structural ambidexterity), the creation of a behavioral context (contextual ambidexterity), or TMT behavioral integration which moderate this relationship. Lin et al. (2007) have also found empirical support for their hypothesis that a firm with a high degree of centrality in the alliance network will tend to have better performance if it adopts an ambidextrous formation of exploratory and exploitative alliances, whereas a focused formation of alliances will tend to bring better performance to firms with a low degree of centrality. This can be explained by the fact that central firms have more ties to other network members, which they can explore and exploit to their benefits (Lin et al., 2007).

Furthermore, network diversity, meaning the number of different social systems that are part of an organization's relationships, might positively influence ambidexterity (Simsek, 2009). A greater amount of novel information, heterogeneity in problem-solving, and greater access to resources can explain this positive relationship (Simsek, 2009). Again, a dual structure, the creation of a behavioral context, or TMT behavioral integration might positively moderate the relationship between network diversity and organizational performance (Simsek, 2009). Simsek (2009) further suggests that environmental complexity positively moderates the relationship between ambidexterity and network centrality and network diversity respectively. An environment is considered complex when an organization has to deal with a variety of heterogeneous actors and activities which are outside its boundaries of the organization's strategic decision-making (Simsek, 2009). This implies that environmental complexity requires higher levels of ambidexterity which can be achieved by network centrality and diversity because they reduce complexity and help to maintain an appropriate level of fit with the environment.

Furthermore, Lin et al. (2007) empirically found that a firm with a high degree of brokerage positions in the interfirm network will tend to have better performance if it adopts a focused formation of either exploratory or exploitative alliances, whereas an ambidextrous formation of alliances will tend to bring better performance for firms with few structural holes. The degree of brokerage positions refers to the ability of a firm to establish relationships with new alliance partners or, in other words, to "span the holes" (Lin et al., 2007). In addition to this, Lin et al. (2007) found empirical support that a firm with an ambidextrous formation of exploratory and exploitative alliances will tend to have better performance in early years of the network, whereas a firm with a focused formation of either exploratory or exploitative alliances will tend to have better performance in later years of the network. This means that when a firm operates in a network context that is rather new, firms have little experience which they can refer to, making a focused approach on either exploration or exploitation rather risky (Lin et al., 2007). Therefore, in a young network, firms will benefit when they take on an ambidextrous design (Lin et al., 2007).

Luo and Rui (2009) investigated the properties and dimensions of emerging market multi-national enterprises' (EM MNEs) ambidexterity. They found that co-orientation, co-competence, co-opetition, and co-evolution determine how good an EM MNE can balance two contrasting elements that occur at the same time. Co-orientation refers to the balance of seeking both short-term survival and long-term growth simultaneously. Co-competence occurs when EM MNEs simultaneously use both transactional and relational capabilities when they operate in international contexts. Co-opetition means that EM MNEs simultaneously compete and cooperate with international stakeholders. Lastly, co-evolution denotes that EM MNEs concurrently respond to and actively influence their external environment in both, their home and host countries. The collective implementation of these sets of systems and processes should help to facilitate ambidexterity in an organization.

Conclusions

The findings obtained from the academic literature reveal that different characteristics of an organization's network influence the strategies for achieving ambidexterity. The results show that managers should implement ambidexterity, meaning the simultaneous pursuit of exploration and exploitation, when their organization has a high degree of network centrality. On the other hand, a one-sided focus on either exploration or exploitation should be implemented when the firm has only low levels of network centrality. Furthermore, the results imply that when an organization has a more diverse network, then this positively influences ambidexterity. In general, network centrality and diversity foster ambidexterity when a firm is confronted with complex environments. The age of the network also influences ambidexterity, so that when firms operate in young and new networks, they benefit from an ambidextrous approach; whereas in older and already established networks, firms can also consider a one-sided focus on either exploration or exploitation. In sum, the results show that managers should orient their ambidexterity-strategy according to the specific characteristics of their organization's network.

4.3.4. Facilitate ambidexterity through externalization and absorptive capacity

Absorptive capacity can help a firm to manage the tensions which arise from simultaneously pursuing exploration and exploitation (Rothaermel and Alexandre, 2009). Absorptive capacity can be defined as a firm's ability to sense, gather and incorporate new knowledge and is enabled by the interaction of exploitative and explorative activities (Andriopoulos and Lewis, 2009). Absorptive capacity plays an important role whenever organizations try to resolve paradoxical tensions concerning the alignment of exploitation and exploration through externalization (Raisch et al., 2009). Externalization can, for example, be conducted in the form of outsourcing or by establishing alliances (Lavie and Rosenkopf, 2006) in order to externalize one or another activity. However, the externalization of either exploration or exploitation is not always easy. Benner and Tushman (2003) criticize that the strategic integration across independently operating firms may be a great challenge. Still, Raisch et al. (2009) argue that externally acquired knowledge may be conducive to the reconfiguration of established knowledge bases, in that ambidexterity does not only require internal and external knowledge processes, but also their integration across the boundaries of an organization.

Linking these findings to the literature on absorptive capacity, Cohen and Levinthal (1990; as cited in Raisch et al., 2009) argue that both external knowledge acquisition and internal knowledge processing are essential, however, there needs to be a balance so that the one is not overwhelmed by the other. Therefore, the ability to integrate internal and external knowledge bases, and thus to become ambidextrous, is largely dependent on a combination of external brokerage and internal absorptive capacity (Raisch et al., 2009). To conclude the findings on the internal-external tension, Rothaermel and Alexandre (2009) argue that ambidexterity in a firm's technology sourcing strategy not only relates to the trade-offs necessary to simultaneously pursue exploration and exploitation, but also form the trade-offs regarding the integration of internal and external technology sourcing. Rothaermel and Alexandre (2009) found empirical support that the absorptive capacity of a firm moderates the inverted U-shaped relationship between a firm's total technology sourcing mix (of existing and new technology) and firm performance so that the positive effect of ambidexterity in technology sourcing on performance is stronger when the firm possesses higher levels of absorptive capacity.

Similarly, Lichtenthaler and Lichtenthaler (2009) propose that the alignment of both internal and external knowledge management processes might offer new avenues in the ambidexterity research, as so far the ambidexterity literature has primarily focused on the internal alignment of exploitation and exploitation (Raisch and Birkinshaw, 2008). More specifically, organizations do not only need to achieve an internal balance, but also have to develop their knowledge processes in that they can compete outside the boundaries of an organization (Lichtenthaler and Lichtenthaler, 2009). The buildup of a knowledge management capacity can help to reconfigure and realign knowledge capacities (Lichtenthaler and Lichtenthaler, 2009).

Conclusions

The findings from the academic literature reveal that the absorptive capacity of an organization, meaning the ability to sense and integrate new knowledge, can help a firm to reconcile exploration and exploitation and to, thus, become ambidextrous. Absorptive capacity especially plays an important role when a firm tries to achieve ambidexterity through the externalization of either explorative or exploitative activities. This means that the ability to absorb and to incorporate the knowledge from the externalized activities is vital for a firm's success. In sum, managers need to develop a knowledge management capacity which enables their organization to not only reconcile exploration and exploitation within their company, but also to integrate the knowledge retrieved from outside the organization.

4.3.5. Facilitate ambidexterity through dynamic capabilities and routines

Dynamic capabilities are one way to cope with paradoxical tensions at the organizational level (Smith and Lewis, 2011) and it is necessary to develop a dynamic capability in order to initiate effective ways of achieving ambidexterity (Jansen et al., 2009). Dynamic capabilities involve processes, routines, and skills (Teece et al., 1997) that allow leaders to respond to changes in the environment which, as a consequence, enables the members of an organization to accept the paradoxical tensions which arise in dynamic environments (Smith and Lewis, 2011). In other words, ambidexterity can be regarded as a dynamic capability in that it involves the routines and processes through which an organization mobilizes, coordinates, and integrates different contradictory agendas and reallocates, combines, and reconfigures different sets of resources across explorative and exploitative units (Teece, 2007; Jansen et al., 2009). Or, as Helfat and Winter state, "ambidexterity relies in part on dynamic capabilities of top managers to perform targeted integration of emerging and mature businesses" (Helfat and Winter, 2011, p. 1248). However, ambidexterity can only become a dynamic capability if it is managed in a repeated fashion and includes the intentional allocation and reconfiguration of firm resources (O'Reilly and Tushman, 2008).

Moreover, as the necessity for exploration and exploitation may differ across different actions and over time, the reconciliation of differentiation and integration tactics may be an important dynamic capability for the development of ambidexterity (Raisch et al., 2009; Gulati and Puranam, 2009). Within independent units in which exploitation and exploration are balanced simultaneously, such as in contextual ambidexterity, meta-routines enable the coordination, synchronization and integration of these two agendas (Simsek et al., 2009; Adler et al., 1999). In other words, routines can help to integrate exploration and exploitation within a single domain (Simsek et al., 2009). Gibson and Birkinshaw (2004)

also highlight the need for meta-routines which should help to systematize the processes which are necessary for aligning exploration and exploitation. Meta-routines can be described as routines for systematically changing existing routines and are responsible for supporting the accomplishment of non-routine activities (Adler et al., 1999).

Findings from practitioner literature

The findings from the practitioner literature also reveal that learning plays an important role when implementing ambidexterity into an organization. Tushman and O'Reilly (1996) state that organizational learning is concerned with incorporating things that go well into an organization and to continually refine the business according to the feedback received from the market. Referring to ambidexterity, this means that its underlying processes are explicitly learned and managed by senior executives (O'Reilly et al., 2009). Ambidexterity therefore can be denoted a dynamic capability, as it involves a set of routines and processes which are repeatable and orchestrated by the top management team (O'Reilly et al., 2009). More specifically, O'Reilly and Tushman describe ambidexterity as a dynamic capability as follows: "As a dynamic capability, ambidexterity embodies a complex set of routines including decentralization, differentiation, targeted integration, and the ability of senior leadership to orchestrate the complex trade-offs that the simultaneous pursuit of exploration and exploitation requires" (O'Reilly and Tushman, 2011, p. 6).

Conclusions

Findings from both, the academic and the practitioner literature, show that an organization's dynamic capabilities can help to cope with the conflicting demands of exploration and exploitation. The academic literature shows that ambidexterity can be regarded as a dynamic capability when managers are able to continually allocate and reconfigure organizational assets in order to integrate emerging and existing businesses. Furthermore, meta-routines, which are part of dynamic capabilities, can help to integrate exploration and exploitation within the same business unit, to change existing routines, and to facilitate the pursuit of nonroutine activities. Findings from the practitioner literature, similarly, show that ambidexterity can be considered as a dynamic capability as it incorporates a set of routines which are responsible for aligning and integrating the differentiated units responsible for exploration and exploitation. Managers can, therefore, use dynamic capabilities in order to better host the contradictory demands of exploration and exploitation and to, consequently, enable their organization to act ambidextrously.

5. Discussion and avenues for future research

This study used a systematic literature review to show which practical implications for managers can be found in academic and practitioner literature to achieve ambidexterity in practice. It was assumed that the management has a vital and overarching function in implementing different mechanisms which should facilitate the achievement of ambidexterity. The results showed that the management should function on three different levels. First, there exist measures which should be implemented directly at the top management level. Second, there are measures concerning the organizational design which should be taken by the management to achieve ambidexterity within the remaining organization. Last, there are moderators and other external factors which are more or less out of scope for decision-making and which can only marginally be influenced by managers. However, the top management can take account of these (sometimes hard to influence) factors and respond accordingly.

5.1. Revision of the implications at the TMT level

The results of this study show that at the top management level, managers need to make important decisions about the strategies and processes which should facilitate ambidexterity. However, this requires them to first be aware of the paradoxical tensions which exist when balancing two contradictory agendas, namely exploration and exploitation (Smith and Tushman, 2005; Smith and Lewis, 2011). In order to resolve these conflicts, managers need to constantly make trade-offs between exploration and exploration and continually allocate and reallocate resources between the two (Smith and Tushman, 2005; O'Reilly and Tushman, 2011). This means that managers not only need to host the inconsistencies that arise from balancing exploration and exploitation, but also recognize and make use of the synergies between them (Smith and Tushman, 2005; Smith and Lewis, 2011).

Paradoxical cognition enables managers to make balanced decisions regarding the achievement of ambidexterity. What such strategic choices should include will be explained here. The results show that various authors (e.g., Gibson and Birkinshaw, 2004; Jansen et al., 2008; Simsek, 2009, O'Reilly and Tushman, 2008) highlight the importance of a shared vision among senior managers, as well as a common culture (Tushman and O'Reilly, 1996) in order to reconcile contradictory agendas. This implies that managers should collaboratively create a vision that emphasizes the need for ambidexterity in an organization. This vision should then be communicated to the other members of the organization in order to create a common an overall culture and identity (O'Reilly and Tushman, 2008, 2011). Shared sets of norms, values, and goals help to align the contrasting elements of exploration and exploitation and, as a consequence, facilitate ambidexterity (Tushman et al., 2011; O'Reilly and Tushman, 2011). In sum, an overarching identity provides for a common strategic intent which aims at aligning separate units for exploration and exploitation which each involve different competencies, processes, systems, and cultures (O'Reilly and Tushman, 2011).

The recognition and resolution of paradoxical tensions and the creation of a common strategic intent can, therefore, be regarded as first steps into directing an organization towards becoming ambidextrous. However, these two steps require that managers display certain behaviors that enable them to become aware of these paradoxical conflicts and to

accordingly take appropriate actions. The results of this thesis show that managers need complex behavioral repertoires which provide them with the flexibility needed to align exploration and exploitation (Smith and Lewis, 2011; Raisch and Birkinshaw, 2008; O'Reilly and Tushman, 2008; Gibson and Birkinshaw, 2004). Furthermore, ambidextrous managers need to be fulfilling multiple roles at a time which are related to the pursuit of both explorative and exploitative activities (Adler et al., 1999; Smith and Lewis, 2011). A managerial mindset helps managers to flexibly respond to changes, to think innovatively, to tolerate failures, and to encourage risk-taking. Especially transformational leaders have the necessary requirements for encouraging critical debate and open discussion about the conflicting demands of exploration and exploitation (Hotho and Champion, 2011; Jansen et al., 2008). Furthermore, transformational leaders have a supportive function and the ability to motivate their followers to aspire goals which are directed towards becoming and acting ambidextrously (Hotho and Champion, 2011; Jansen et al., 2008; Tushman and O'Reilly, 1996).

In order to manage the paradoxical tensions regarding exploration and exploitation, results from the academic as well as the practitioner literature show that there also are certain characteristics and constellations of top management teams which should facilitate this task. First of all, a team which is behaviorally (Lubatkin et al., 2006) or socially (Jansen et al., 2009) integrated is better able achieve ambidexterity in an organization. Behavioral integration allows team members to collaboratively work together and to exchange knowledge and information which gives them the possibility to resolve conflicts and to openly discuss new ideas and opportunities (Lubatkin et al., 2006; Jansen et al., 2009). Furthermore, the affiliations to prior companies play an important role in how managers cope with paradox (Beckman, 2006). The results show that top management teams whose members have both common and diverse prior company affiliations are better able to develop a common understanding and to transfer knowledge (Beckman, 2006). Common prior company affiliations foster a common culture and mutual trust and understanding among team members which are beneficial to exploitation (Beckman, 2006). Diverse prior company affiliations promote broader access to knowledge and networks and are therefore beneficial to exploration (Beckman, 2006). Lastly, both academic and practitioner literature reveal that there are two specific structural constellations of teams which should facilitate the integration of strategic contradictions regarding exploration and exploitation. The first constellation suggests that teams should be leader-centered, meaning that the leader has the responsibility to integrate exploration and exploitation by relying on heavy exchange with the other team members (Smith and Tushman, 2005; Tushman et al., 2011). The second constellation suggests that teams should be team-centered, meaning that the members collaboratively work together in order to integrate contradictory agendas (Smith and Tushman, 2005; Tushman et al., 2011). Both possibilities seem to be workable solutions and every organization needs to decide for themselves which approach they want to adopt. Furthermore, the results show that managers can use formal and informal coordination mechanisms to facilitate ambidexterity. The two main formal structural mechanisms include a manager's level of decision-making authority (Raisch et al., 2009; Mom et al., 2009) and the formalization of tasks in a business unit (Mom et al., 2009; Jansen et al., 2009). The findings show that the higher a manager's decision-making authority, the better is he or she able to integrate explorative and exploitative activities (Raisch et al., 2009; Mom et al., 2009). Pertaining to formalization, the results suggest that the formalization of tasks enhances the level of exploitative behavior within a unit (Jansen et al., 2009). Both, a manger's level of decision-making authority and the formalization of tasks, positively influence ambidexterity. Informal, or personal, coordination mechanisms include a manager's participation in cross-functional interfaces and his or her connectedness to other members of the organization (Mom et al., 2009; Adler et al., 1999; Jansen et al., 2009). Both mechanisms do not only facilitate the exchange of knowledge between organizational members, but also increase the trust between a manager and others (Mom et al., 2009; Jansen et al., 2009). This, in turn, leads to the ability to better resolve paradoxical conflicts and to achieve ambidexterity (Mom et al., 2009; Jansen et al., 2009). Moreover, the results suggest that formal and informal coordination mechanisms positively correlate with each other. The findings show that a manager's decision-making authority positively interacts with the participation in cross-functional interfaces and the connectedness to other members of the organization (Mom et al., 2009). For managers, this means that when they succeed in having authority over decisions and when they deploy dense networks of direct personal contacts to other organizational members, they are more capable of acting ambidextrous themselves in order to implement measures to direct their organization towards ambidexterity (Mom et al., 2009; Jansen et al., 2009).

5.1.1. Avenues for future research

The revision of the results presented above give rise to multiple avenues for future research. First of all, it would be interesting to more specifically distinguish between the individual manager and the top management team as a whole. One leverage point could be, for instance, the more in depth analysis of a manager's individual behavior which enables him to act ambidextrously. Another leverage point could concern the dynamic and the culture of a group of senior leaders. Pertaining to this, it would be interesting to examine which factors influence the top management team to collaboratively work together to reconcile the contrasting demands of exploration and exploitation. Second, it might be interesting to also examine the role of lower level managers as the focus of prior research has primarily lied on the role of the top management team. Certainly, managers at lower hierarchical levels have the responsibility to implement the strategic decisions made at the top management level. Therefore, a deeper understanding of the functions and characteristics of the managers at the operating levels would lead to greater

insights into how ambidexterity can be achieved in practice.

5.2. Revision of the implications concerning organizational design

The measures which can be implemented directly at the top management team level serve as important prerequisites for applying organizational design measures within the remaining organization. The two main solutions in this regard include structural and contextual arrangements. In addition to this, human resource practices have a supportive function in implementing these structural and contextual solutions in order to achieve ambidexterity within the organization. This section has the aim to revisit the core findings concerning organizational design mechanisms, to highlight their importance for managers, and to provide avenues for future research pertaining to this subject.

As already mentioned earlier, structural ambidexterity involves the spatial separation of distinct and autonomous units responsible for either exploration or exploitation (Benner and Tushman, 2003; Simsek, 2009; Jansen et al., 2009; Tushman and O'Reilly, 1996). The structural separation of the units avoids that the processes, structure, and cultures of one unit are not overwhelmed by the other (Simsek, 2009). Following Benner and Tushman (2003) suggestion, then the tasks, the required competencies, the culture, and the structural arrangements (O'Reilly and Tushman, 2004) are constant, and therefore tightly coupled, within the subunit. However, across the different subunits responsible for either exploration or exploitation, these different elements are not consistent and only loosely coupled with each other (Benner and Tushman, 2003). This leads to the conclusion that each unit operates independently within structural ambidexterity. Still, these differentiated subunits are organizationally interdependent which requires their integration at the top management team level (Benner and Tushman, 2003; Simsek et al., 2009). For managers this, again, means to host the contradictions that arise from the deployment of differentiated units responsible for exploration and exploitation and to integrate them with regard to the vision, the goals, and the culture of the organization.

The second main mechanism for achieving ambidexterity is the creation of a context that allows the simultaneous pursuit of exploration and exploitation within the same business unit (Birkinshaw and Gibson, 2004; Gibson and Birkinshaw, 2004; Simsek et al., 2009). With regard to contextual ambidexterity, the implication for managers is to create a context which equally emphasizes high performance, through a combination of discipline and stretch, and social support, through a combination of trust and support (Gibson and Birkinshaw, 2004). Additionally, managers should be aware that contextual ambidexterity only works, when the members of the organization act and behave ambidextrously (Birkinshaw and Gibson, 2004). The ability to act ambidextrously gives individuals the capability to decide on how they want to divide their time between explorative and exploitative activities (Gibson and Birkinshaw, 2004).

Furthermore, the findings of the systematic review of the academic literature reveals that there are a variety of human resource practices which facilitate the achievement of ambidexterity. Specifically, these include certain job designs, such as job enrichment or task partitioning (Adler et al., 1999), and the training of the employees to enable them to work efficiently in an ambidextrous organization (Simsek et al., 2009). In addition to this, rewards and other incentives can enhance the ambidexterity among employees (Jansen et al., 2008; Jansen et al., 2009; Kaplan and Henderson, 2005). Especially the top management team benefits from senior team contingency rewards which have the aim to increase trust and collaboration (Jansen et al., 2008; Jansen et al., 2009; Smith and Tushman, 2005). In general, managers should advocate for intrinsically motivating incentives which should foster the creation of knowledge and provide a sense of ownership of the work (Hotho and Champion, 2011). Moreover, managers should be aware that with the development and the efficient management of intellectual capital within their organization, they can effectively direct the adoption of explorative or exploitative activities through either human, social, or organizational capital (Kang and Snell, 2009).

5.2.1. Avenues for future research

Structural and contextual ambidexterity have been widely discussed as antecedents of ambidexterity by researches. What would be an interesting avenue for research is to examine structural and contextual as complementary solutions for achieving ambidexterity (see Birkinshaw and Gibson, 2004). The underlying assumption here is that the separate units responsible for either explorative or exploitative activities might not be well connected to the core business (Birkinshaw and Gibson, 2004). Viewing contextual ambidexterity as a complement to structural ambidexterity might therefore lead to new leverage points regarding the antecedents of ambidexterity. In addition to this, there is relatively little knowledge regarding the human resource practices which can facilitate the implementation and the development of ambidexterity. Therefore, future research would benefit from a more in depth examination of other possible HR practices which might foster ambidexterity. Although some researchers (e.g., Simsek et al., 2009; Adler et al., 1999) already approached the necessity of training and job enrichment programs to enable employees to perform both exploration and exploitation, it has not yet been examined which other job designs could possibly influence the ambidexterity of employees. For instance, it could be tested whether job rotation or job enlargement have an effect on employee's ambidexterity. Both of these job designs are directed towards motivating employees to extend their activities and capabilities. Consequently, job rotation and job enlargement can be regarded as an important HR practices for achieving ambidexterity because they not only require the creation of new knowledge and the application of existing knowledge, but also the recognition and resolution of paradoxes.

5.3. Revision of the external factors and moderators

The findings of this thesis show that some factors which influence ambidexterity can only be marginally be influenced by managers. These include the amount of resources that a firm possesses, certain environmental factors, different characteristics of the organization's network, the organization's absorptive capacity and its dynamic capabilities. However, managers can build awareness of these particularities and orient their ambidexterity strategy according to the special requirements of each of these factors. This section will critically revisit the single results obtained from the systematic literature review.

First of all, the findings pertaining to the availability of resources have specific implications regarding the strategy of ambidexterity which managers should deploy. More precisely, this means that managers should rather use a structurally ambidextrous design when their firm possesses high levels of organizational slack and a contextually ambidextrous design when their firms a rather small and do not possess sufficient resources to support structurally separate units for exploration and exploitation (Voss et al., 2008; Lubatkin et al., 2006, Raisch and Birkinshaw, 2008; Lin et al., 2007). Second, the results suggest that ambidexterity has a positive influence on a firm's performance when this firm operates in a complex or dynamic environment (Simsek, 2009, Jansen et al., 2006). Managers also need to be aware that environmental dynamism has a positive moderating effect on the relationship between exploration and financial performance (Jansen et al., 2006). This also implies that managers have to sense changes in new a dynamic markets through exploration and to foster exploitation in existing markets (Eisenhardt et al., 2010). Third, managers should consider the different characteristics of their organizational network when they implement an ambidexterity-oriented strategy. The results show that when an organization is central in a network, then managers should use an ambidextrous approach to fully reap the benefits arising from the connections to others in this network (Lin et al., 2007). Furthermore, the results indicate that the more diverse the network of an organization is, the more positive its effect on ambidexterity (Lin et al., 2007). Last, the findings imply that ambidexterity is beneficial to a firm when it operates in a rather young and new network (Lin et al., 2007). In the case of multinational enterprises in emerging markets, for example, co-orientation, co-competence, co-evolution, and co-opetition seem to be vital dimensions of ambidexterity through which organizations can establish their position in new networks (Luo and Rui, 2009). In sum, the results obtained from the academic literature suggest that the availability of resources, various environmental factors, as well as certain characteristics of an organization's network need to be considered by managers in order to direct their ambidexterity strategy towards the emerging needs of each of these particularities.

In addition to, the results suggest that organizations can make use of their absorptive capacity and their dynamic capabilities in order to facilitate and to support ambidexterity. Pertaining to absorptive capacity, for managers this implies that they could consider the externalization of single activities of exploration or exploitation (Raisch et al., 2009; Lavie and Rosenkopf, 2006). In such a case, the absorptive capacity of an organization can be used to integrate the externally acquired knowledge and to reconfigure already established knowledge bases (Raisch and Birkinshaw, 2008; Raisch et al., 2009; Lichtenthaler and Lichtenthaler, 2009). Furthermore, managers can make use of the dynamic capabilities of their firm in that they help to continually allocate and reconfigure organizational assets to integrate new and existing businesses (Smith and Lewis, 2011; O'Reilly and Tushman, 2008; O'Reilly et al., 2009). Meta-routines, which are part of dynamic capabilities, can also be implemented by managers in order to enable employees to host the demands of both, explorative and exploitative activities (Simsek et al., 2009; Adler et al., 1999). In sum, the results obtained from the systematic review of the selected literature reveal that the absorptive capacity and the dynamic capabilities of an organization represent important facilitators of ambidexterity which should be considered by managers who want to achieve ambidexterity within their organization.

5.3.1. Avenues for future research

The effect of the availability of resources and the environmental factors on a firm's ambidexterity and its performance have already been examined empirically by a variety of researchers (e.g., Lubatkin et al., 2006; Jansen et al., 2006). Although managers only have a limited scope for decision-making regarding these aspects, one avenue for future research could be to analyze how managers can react when there are sudden changes in one of these domains. For instance, when a new and dynamic market, which emerges from the launch of a new innovation, turns into a stable and mature market, then it would be interesting to analyze how managers can use ambidexterity to adapt to these changes. Specifically, it would be important to consider the processes by which managers and top management teams sense changes in their environments and how ambidexterity can help them to respond to these changes in an efficient way. A second possible avenue for future research concerns the dynamic capabilities of a firm. O'Reilly and Tushman (2008) argue that dynamic capabilities need to be repeatable in order to be useful for managers and the entire organization. This implies that ambidexterity should also be a repeatable and controllable processes which not only emerges "contingently" through either external or internal changes within an organization. Therefore, it would be interesting to further examine how managers and top management teams can create sustainable routines for constantly aligning and reallocating resource assets that should help to reconcile exploration and exploitation and to, consequently, facilitate ambidexterity.

6. Conclusion

This study's aim was to derive practical implications for managers from academic articles and practitioner literature

concerning the concept of ambidexterity. The results obtained from the systematic literature review provide some basic principles which should help to achieve ambidexterity in an organization and to bridge the rigor-relevance regarding this topic. This section will summarize the main guidelines for managers regarding the achievement of ambidexterity. Additionally, some limitations of this study will be explained and it will be highlighted if and to which extent this study helped in bridging the rigor-relevance gap regarding ambidexterity.

6.1. Summary of results

There are some major prerequisites which are necessary for implementing ambidexterity in an organization. The starting point here is the top management which has a vital function in reconciling the contrasting and often conflicting demands of exploration and exploitation in order to achieve ambidexterity. First of all, managers need to recognize and resolve the paradoxical tensions that arise from these two contrasting agendas. Second, managers need to develop an ambidexterity-oriented strategy and communicate this strategy across the members of the whole organization. These processes can be facilitated by ambidextrous leaders with complex behavioral repertoires and specific constellations and characteristics of top management teams. Thirdly, managers should make use of both formal and informal (or personal) coordination mechanisms to foster not only their own ambidexterity, but also the ambidexterity within the remaining organization.

Once the management has succeeded in implementing the necessary measures at the top management team level, the rest of the organization can now be designed to act ambidextrously. Specifically, managers can use either structural or contextual mechanisms to do so. Structural solutions include the creation of spatially separated subunits which are each responsible for either exploration or exploitation. Contextual solutions involve the creation of a context which allows individuals to simultaneously pursue exploration and exploitation within the same unit. In addition to this, specific human resource practices have a supportive function in fostering ambidexterity among the members of an organizations. These include primarily ambidexterity-oriented incentives, job enrichment and training, and the development of human, social, and organizational capital.

Lastly, managers need to respond to certain external factors and other moderators which might influence the achievement of ambidexterity. First of all, managers need to orient their strategy for achieving ambidexterity in consideration of the availability of the resources that the organization has. Second, managers have to be aware of certain environmental influences and take appropriate measures as the external environment requires. Third, certain characteristics of an organization's network can be used by managers to implement according actions for the reconciliation of exploration and exploitation. Fourth, organizations can make use of their absorptive capacity and their dynamic capabilities in order to further foster their ambidexterity.

6.2. Limitations

The implications and results of this study are subject to several limitations. One possible limitation concerns the methodological proceeding of this thesis. The 40 most cited articles treating the topic of ambidexterity were used to filter out practical implications for managers. Although this method contributed to selecting the most influential studies concerning ambidexterity, it was limited in that it did not include more recent articles. The 40 most cited articles mainly included papers from the years 2004 to 2009. Therefore, more recent studies, meaning those published after 2009 were, apart from a few exceptions, neglected. Due to the great proliferation of articles concerning ambidexterity in the last few years, it would therefore be interesting to also examine more recent articles with regard to this thesis' research question.

A second limitation is that it is unclear whether the practical implications derived from the academic and practitioner literature are taken up in practice or not. Although there are certain approaches which deliver already very concrete implications for managers on how to best achieve ambidexterity in their organization, it remains uncertain if these implications are really being implemented in practice. One reason for this is that although research on ambidexterity has largely increased over the last years, the construct of ambidexterity still lacks a concrete conceptualization and theorization. Although this study refers to different disciplines and theories of ambidexterity in order to integrate and extend the current understanding, many aspects of this construct remain unexplained and not understood. Therefore, more research regarding ambidexterity and the development of possibly common theories and conceptualizations of this construct would help to further explore the practical implications of ambidexterity.

With regard to the results obtained from the systematic review of articles from academic and from practitioner literature, it can be said that this thesis partly succeeded to bridge the rigor-relevance gap concerning the topic of ambidexterity. The results provide general principles for managers to achieve ambidexterity in an organization. However, the abstract nature of the construct of ambidexterity makes it difficult to define common conceptualizations and to operationalize this concept in practice. What might be useful in this regard is, for instance, the distinction into either abstract or concrete practical implications. Concrete solutions would then provide managers with very explicit and detailed implications on how to achieve ambidexterity in practice, which cannot be derived from abstract solutions. More concrete solutions would probably improve the extent to which the rigor-relevance gap is bridged and enable the deduction of more consistent results with regard to ambidexterity.

References

Adler, P., Goldoftas, B., and Levine, D. I. Flexibility Versus Efficiency? A Case Study of Model Changeovers in the Toyota Production System. *Organization Science*, 10(1):43–68, 1999.

Ambos, T., Mäkelä, K., Birkinshaw, J., and D'Este, P. When Does University Research Get Commercialized? Creating Ambidexterity in Research Institutions. *Journal of Management Studies*, 45(8):1424–1447, 2008.

Andriopoulos, C. and Lewis, M. Exploitation-Exploration Tensions and Organizational Ambidexterity: Managing Paradoxes of Innovation. *Organization Science*, 20(4):696–717, 2009.

Beckman, C. The Influence of Founding Team Company Affiliations on Firm Behavior. *Academy of Management Journal*, 49(4):741–758, 2006.

Benner, M. and Tushman, M. Exploitation, Exploration, and Process Management: The Productivity Dilemma Revisited. *Academy of Management Review*, 28(2):238–256, 2003.

Birkinshaw, J. and Gibson, C. Building Ambidexterity Into an Organization. *MIT Sloan Management Review*, 45(4):47–55, 2004.

Birkinshaw, J. and Gupta, A. Clarifying the Distinctive Contribution of Ambidexterity to the Field of Organization Studies. *Academy of Management Perspectives*, 27(4):287–298, 2013.

Burgelman, R. Intraorganizational Ecology of Strategy Making and Organizational Adaptation: Theory and Field Ressearch. *Organization Science*, 2(3):239–262, 1991.

Burgelman, R. Strategy as vector and the inertia of coevolutionary lock-in. *Administrative Science Quarterly*, 47(2):325–357, 2002.

Cao, Q., Gedajlovic, E., and Zhang, H. Unpacking Organizational Ambidexterity: Dimensions, Contingencies, and Synergistic Effects. *Organization Science*, 20(4):781–796, 2009.

Dess, G. and Lumpkin, G. The Role of Entrepreneurial Orientation in Stimulating Effective Corporate Entrepreneurship. *Academy of Management Perspectives*, 19(1):147–156, 2005.

Eisenhardt, K., Furr, N., and Bingham, C. Microfoundations of Performance: Balancing Efficiency and Flexibility in Dynamic Environments. *Organization Science*, 21(6):1263–1273, 2010.

Floyd, S. and Lane, P. Strategizing Throughout the Organization: Managing Role Conflict in Strategical Renewal. *Academy of Management Review*, 25 (1):154–177, 2000.

Gibson, C. and Birkinshaw, J. The Antecedents, Consequences, and Mediating Role of Organizational Ambidexterity. *Academy of Management Journal*, 47(3):209–226, 2004.

Gulati, R. and Puranam, P. Renewal Through Reorganization: The Value of Inconsistencies Between Formal and Informal Organization. *Organization Science*, 20(2):422–440, 2009.

Gupta, A., Smith, K., and Shalley, C. The Interplay Between Exploration and Exploitation. *Academy of Management Journal*, 49(4):693–706, 2006.

Güttel, W. and Konlechner, S. Continuously Hanging by a Thread: Managing Contextually Ambidextrous Organizations. *Schmalenbach Business Review*, 61(2):149–171, 2009.

He, Z. and Wong, P. Exploration vs. Exploitation: An Empirical Test of the Ambidexterity Hypothesis. *Organization Science*, 15(4):481–494, 2004.

Helfat, C. and Winter, S. Untangling Dynamic and Operational Capabilities: Strategy for the (N)ever-Changing World. *Strategic Management Journal*, 32(11):1243–1250, 2011.

Hodgkinson, G. and Rousseau, D. Bridging the Rigour-Relevance Gap in Management Research: It's Already Happening! *Journal of Management Studies*, 46(3):534–546, 2009.

Hodgkinson, G., Herriot, P., and Anderson, N. Record Re-aligning the Stakeholders in Management Research: Lessons from Industrial, Work and Organizational Psychology. *British Journal of Management*, 12(Peer Reviewed):S41–S48, 2001.

Hotho, S. and Champion, K. Small businesses in the new creative industries: innovation as a people management challenge. *Management Decision*, 49 (1):29–54, 2011.

Im, G. and Rai, A. Knowledge Sharing Ambidexterity in Long-Term Interorganizational Relationships. *Management Science*, 54(7):1281–1296, 2008.

Jansen, J., van den Bosch, F., and Volberda, H. Exploratory Innovation, Exploitative Innovation, and Performance: Effects of Organizational Antecedents and Environmental Moderators. *Management Science*, 52(11): 1661–1674, 2006.

Jansen, J., George, G., van den Bosch, F., and Volberda, H. Senior Team Attributes and Organizational Ambidexterity: The Moderating Role of Transformational Leadership. *Journal of Management Studies*, 45(5): 982–1007, 2008.

Jansen, J., Tempelaar, M., van den Bosch, F., and Volberda, H. Structural Differentiation and Ambidexterity: The Mediating Role of Integration Mechanisms. *Organization Science*, 20(4):797–811, 2009.

Junni, P., Sarala, R., Taras, V., and Tarba, S. Organizational Ambidexterity and Performance: A Meta-Analysis. *Academy of Management Perspectives*, 27(4):299–312, 2013.

Kang, S. and Snell, S. Intellectual Capital Architectures and Ambidextrous Learning: A Framework for Human Resource Management. *Journal of Management Studies*, 46(1):65–92, 2009.

Kaplan, S. and Henderson, R. Inertia and Incentives: Bridging Organizational Economics and Organizational Theory. *Organization Science*, 16 (5):509–521, 2005.

Kieser, A. and Leiner, L. Why the Rigour-Relevance Gap in Management Research Is Unbridgeable. *Journal of Management Studies*, 46(3):516–533, 2009.

Lavie, D. and Rosenkopf, L. Balancing Exploration and Exploitation in Alliance Formation. *Academy of Management Journal*, 49(4):797–818, 2006.

Lichtenthaler, U. and Lichtenthaler, E. A Capability-Based Framework for Open Innovation: Complementing Absorptive Capacity. *Journal of Management Studies*, 46(8):1315–1338, 2009.

Lin, Z., Yang, H., and Demirkan, I. The Performance Consequences of Ambidexterity in Strategic Alliance Formations: Empirical Investigation and Computational Theorizing. *Management Science*, 53(10):1645–1658, 2007.

Lubatkin, M., Simsek, Z., Ling, Y., and Veiga, J. Ambidexterity and Performance in Small-to Medium-Sized Firms: The Pivotal Role of Top Management Team Behavioral Integration. *Journal of Management*, 32(5): 646–672, 2006.

Luo, Y. and Rui, H. An Ambidexterity Perspective Toward Multinational Enterprises From Emerging Economies. *Academy of Management Perspectives*, 23(4):49–70, 2009.

March, J. Exploration and exploitation in organizational learning. *Organization Science*, 2(1):71–87, 1991.

Mayring, P. *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. Beltz Pädagogik, 2015.

McDonough, E. and Leifer, R. Using Simultaneous Structures to Cope with Uncertainty. *Academy of Management Journal*, 26(4):727–735, 1983.

Mom, T., van den Bosch, F., and Volberda, H. Understanding Variation in Managers' Ambidexterity: Investigating Direct and Interaction Effects of Formal Structural and Personal Coordination Mechanisms. *Organization Science*, 20(4):812–828, 2009.

O'Connor, G. and DeMartino, R. Organizing for Radical Innovation: An Exploratory Study of the Structural Aspects of RI Management Systems in Large Established Firms. *The Journal of Product Innovation Management*, 23(6):475–497, 2006.

O'Reilly, C. and Tushman, M. The Ambidextrous Organization. *Harvard Business Review*, 82(4):74–81, 2004.

O'Reilly, C. and Tushman, M. Ambidexterity as a dynamic capability: Resolving the innovator's dilemma. *Research in Organizational Behavior*, 28 (Peer Reviewed):185–206, 2008.

O'Reilly, C. and Tushman, M. Organizational Ambidexterity in Action - How Managers Explore and Exploit. *California Management Review*, 53(4), 2011.

O'Reilly, C. and Tushman, M. Organizational Ambidexterity: Past, Present, and Future. *Academy of Management Perspectives*, 27(4):324–338, 2013.

O'Reilly, C., Tushman, M., and Harreld, J. Organizational ambidexterity. IBM and emerging business opportunities. *California Management Review*, 51 (4):75–99, 2009.

Raisch, S. and Birkinshaw, J. Organizational Ambidexterity: Antecedents, Outcomes, and Moderators. *Journal of Management*, 34(3):375–409, 2008.

Raisch, S., Birkinshaw, J., Probst, G., and Tushman, M. Organizational Ambidexterity: Balancing Exploitation and Exploration for Sustained Performance. *Organization Science*, 20(4):685–695, 2009.

Rosing, K., Frese, M., and Bausch, A. Explaining the heterogeneity of the leadership-innovation relationship: Ambidextrous leadership. *The Leadership Quarterly*, 22(5):956–974, 2011.

Rothaermel, F. and Alexandre, M. Ambidexterity in Technology Sourcing: The Moderating Role of Absorptive Capacity. *Organization Science*, 20 (4):759–780, 2009.

Shapiro, D., Kirkman, B., and Courtney, H. Perceived Causes and Solutions of the Translation Problem in Management Research. *Academy of Management Journal*, 50(2):249–266, 2007.

Sheremata, W. Centrifugal and Centripetal Forces in Radical New Product Development Under Time Pressure. *Academy of Management Review*, 25 (2):389–408, 2000.

Sidhu, J., Commandeur, H., and Volberda, H. The Multifaceted Nature of Exploration and Exploitation: Value of Supply, Demand, and Spatial Search for Innovation. *Organization Science*, 18(1):20–38, 2007.

Simsek, Z. Organizational Ambidexterity: Towards a Multilevel Understanding. *Journal of Management Studies*, 46(4):597–624, 2009.

Simsek, Z., Heavey, C., Veiga, J., and Souder, D. A Typology for Aligning Organizational Ambidexterity's Conceptualizations, Antecedents, and Outcomes. *Journal of Management Studies*, 46(5):864–894, 2009.

Smith, W. and Lewis, M. Toward a Theory of Paradox: A Dynamic Equilibrium Model of Organizing. *Academy of Management Review*, 36(2): 381–403, 2011.

Smith, W. and Tushman, M. Managing Strategic Contradictions: A Top Management Model for Managing Innovation Streams. *Organization Science*, 16(5):522–536, 2005.

Taylor, A. and Helfat, C. Organizational Linkages for Surviving Technological Change: Complementary Assets, Middle Management, and Ambidexterity. *Organization Science*, 20(4):718–739, 2009.

Teece, D. Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance. *Strategic Management Journal*, 28(13):1319–1350, 2007.

Teece, D., Pisano, G., and Shuen, A. Dynamic Capabilities and Strategic Management. *Strategic Management Journal*, 18(7):509–533, 1997.

Tiwana, A. Do bridging ties complement strong ties? An empirical examination of alliance ambidexterity. *Strategic Management Journal*, 29(3): 251–272, 2008.

Tranfield, D., Denyer, D., and Smart, P. Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review. *British Journal of Management*, 14(3):207–222, 2003.

Turner, N., Swart, J., and Maylor, H. Mechanisms for Managing Ambidexterity: A Review and Research Agenda. *International Journal of Management Reviews*, 15(3):317–332, 2013.

Tushman, M. and O'Reilly, C. Ambidextrous Organizations - Managing Evolutionary and Revolutionary Change. *California Management Review*, 38 (4):8–30, 1996.

Tushman, M., Smith, W., and Binns, A. The Ambidextrous CEO. *Harvard Business Review*, 89(6):74–80, 2011.

Uotila, J., Maula, M., Keil, T., and Zahra, S. Exploration, Exploitation, and Financial Performance: Analysis of S&P 500 Corporations. *Strategic Management Journal*, 30(2):221–231, 2009.

Volberda, H., Baden-Fuller, C., and van den Bosch, F. Mastering Strategic Renewal: Mobilising Renewal Jouneys in Multi-unit Firms. *Long Range Planning*, 34(2):159–178, 2001.

Voss, G., Sirdeshmukh, D., and Voss, Z. The Effects of Slack Resources and Environmental Threat on Product Exploration and Exploitation. *Academy of Management Journal*, 51(1):147–164, 2008.

Junior Management Science 1 (2016) 1-300

Junior Management Science

journal homepage: www.jums.academy

Inkommensurabilität – Implikationen für die evolutionäre Organisationstheorie unter begriffstheoretischer und epistemologischer Perspektive

Nicolas Wüthrich
Universität Zürich

Abstract

Innerhalb der Organisationstheorie ist ein Theoriepluralismus zu konstatieren, der sich in einer Vielzahl unterschiedlicher Erklärungsansätze und Konzeptualisierungen des Verhaltens betriebswirtschaftlich organisierter Unternehmen äussert. Dreh- und Angelpunkt für die Beantwortung der Frage, wie dieser Theoriepluralismus zu bewerten ist, bildet der Begriff der Inkommensurabilität. Die vorliegende Untersuchung nimmt sich dieser Thematik an. In einem ersten Schritt wird der Begriff der Inkommensurabilität analysiert. Zielpunkt dieser Begriffsklärung ist ein metatheoretischer Bezugsrahmen, der mit einer begriffstheoretischen und einer epistemologischen Bedingung ein Kriterienpaket bereithält, welches die Grundlagen der Inkommensurabilität sicht- und handhabbar macht. Anhand dieses Bezugsrahmens werden in einem zweiten Schritt exemplarisch Implikationen der Inkommensurabilitätsthematik für die evolutionäre Organisationstheorie dargelegt. Die Implikationen werden in Form von acht inter- und intratheoretischen Befunden erläutert.

Diese Argumentationslinie ermöglicht abschliessend eine reflektierte Einschätzung des Theoriepluralismus und die Exponierung einer neuen Forschungsstrategie für die Grundlagendebatte in der Organisationstheorie.

Keywords: Theoriepluralismus, Inkommensurabilität, Thomas S. Kuhn, Evolutionäre Organisationstheorie

1. Einleitung

Betrachtet man das Feld desjenigen Teils der Organisationstheorie, der sich mit betriebswirtschaftlich organisierten Unternehmen befasst, so fällt Eines unmittelbar auf: Das Feld wird von einer Vielzahl verschiedenster Theorieansätze bevölkert. Exemplarisch können an dieser Stelle ein institutionsökonomischer (Williamson, 1993), ein systemtheoretisch-konstruktivistischer (Luhmann, 1984; Simon, 2007), ein populationsökologischer (Hannan und Freeman, 1989) sowie ein Human-Relations Ansatz (Roethlisberger und Dickson, 1964) genannt werden. Diese vier Ansätze bieten für ihren übereinstimmenden Gegenstandsbereich – Organisationen – unterschiedliche begriffliche Konzeptualisierungen, heterogene Erklärungsansätze sowie verschiedenste Vorhersagen in Bezug auf organisationale Entwicklungspfade an. Illustrativ hierfür sind die unterschiedlichen Organisationsbegriffe im institutionsökonomischen und populationsökologischen Ansatz. Gemäss Williamson (1993, S. 91) ist eine Organisation als homogene Bündelung von Transaktionen zu begreifen, wobei die Existenzberechtigung der Organisation in einem Transaktionskostenvorteil gegenüber der Koordinationsform des Marktes besteht. Demgegenüber konzipieren Hannan und Freeman (1989, S. xii, 15) Organisationen als heterogene soziale Gebilde, die Teil einer Population von Organisationen und somit exogen iniziierten, evolutionärdynamischen Prozessen unterworfen sind.

Dieser Theoriepluralismus ist intuitiv betrachtet ein irritierendes Phänomen. Die erwähnten Ansätze beschäftigen sich als organisationstheoretische Positionen per definitionem mit dem gleichen Gegenstandsbereich. Das Erkenntnisobjekt der verschiedenen Untersuchungen stimmt dementsprechend überein. Damit ist die alltägliche Erwartung verknüpft, dass – zumindest über den Zeitverlauf betrachtet – eine Konvergenz im Hinblick auf die Konzeptualisierungen, Erklärungen sowie Vorhersageleistungen auftreten. Der Umstand, dass dieser Theoriepluralismus ein erläuterungswürdiges Phänomen ist, wird in der Literatur breit anerkannt (vgl. exemplarisch Scherer und Steinmann, 1999, S. 519, 521 f.). Nachfolgend werden zwei miteinander verbundene Reflexionspfade beleuchtet, die sich dazu in der Literatur finden lassen.

Eine erste Bemühung diesen Theoriepluralismus bes-

DOI: http://dx.doi.org/10.5282/jums/v1i1pp188-215

ser fassen zu können, besteht darin, ein Ordnungsschema für die organisationstheoretischen Ansätze zu formulieren. Folgt man dem Strukturierungsvorschlag von Burrell und Morgan (1979) dann lässt sich die Vielzahl der Ansätze anhand zwei metatheoretischer Dimensionen kategorisieren. Sozialwissenschaftlichen Theorien (und damit auch Organisationstheorien) liegen implizit stets wissenschafts- und gesellschafstheoretische Annahmen zu Grunde (vgl. Burrell und Morgan, 1979, S. 1; Burrell, 1997, S. 38 ff.). Gemäss Burrell und Morgan (1979, S.3) lässt sich ein subjektivistisches von einem objektivistischen Wissenschaftsverständnis unterscheiden.[1] Die gesellschaftstheoretischen Annahmen untergliedern sich in ein Verständnis von Gesellschaft als geordnetes, stabiles Ganzes im Gegensatz zu einem Gesellschaftsbild, das auf Diskontinuitäten und Umbrüche fokussiert (vgl. ebd., S. 13, 16 ff.). Die so entstehende Matrix erlaubt die Identifizierung von vier Paradigmen, die in Bezug auf die organisationstheoretischen Ansätze eine strukturierende Funktion haben (vgl. ebd., S. 21 ff.).[2]

Ausgehend von derartigen Strukturierungsversuchen entwickelte sich in einem zweiten Reflexionspfad eine kontroverse Debatte darüber, wie der offenkundige Theoriepluralismus zu bewerten ist. Befürworter der Perspektivenvielfalt (Burrell und Morgan, 1979, S. ix, 398 f.; Burrell, 1996, S. 648; Morgan, 1980, S. 605, 620; Morgan, 1990, S.13, 26 ff.; Hassard, 1993, S. 88 f.; Kirsch et al., 2007, S. 127, 158 f.; Astley und van de Ven, 1983, S. 245 f.; Glynn et al., 2000, S. 731 f.) konstatieren, dass sich nur durch das Eröffnen und Erarbeiten divergenter organisationstheoretischer Ansätze das vielschichtige Phänomen „Organisation“ adäquat erfassen lässt.

Gegen diese Bejahung der Vielfalt wurden primär drei Argumente hervorgebracht. Nicolai und Seidl (im Druck, S. 22 f.) sowie Scherer und Dowling (1995, S. 196) betonen erstens, dass der Theoriepluralismus bei der Anwendung betriebswirtschaftlichen Wissens zu instrumentellen und legitimatorischen Schwierigkeiten führt. Pfeffer (1993, S. 608, 611 f.) hebt zweitens die Notwendigkeit des Konsens hervor, um überhaupt eine produktive Wissensvermehrung in Bezug auf organisationales Verhalten zu erreichen. Schliesslich hält Donaldson (1985, S. 27 ff., 40 ff.) drittens fest, dass verschiedene Ansätze schlicht und einfach überflüssig sind, da der strukturfunktionalistische Ansatz eine umfassende Erklärung organisationaler Phänomene leisten kann. Im Zentrum dieser drei unterschiedlich gelagerten Kritikpunkte steht die Thematik der Inkommensurabilität verschiedener organisationstheoretischer Forschungsansätze (vgl. Tadajewski, 2009, S. 476). Der Inkommensurabilitätsbegriff in dem hier zu behandelnden Sinne wurde im Jahre 1962 simultan durch Kuhn und Feyerabend in die Diskussion eingeführt (vgl. Sankey, 1994, S. 2).[3] Speziell aufgrund der enormen Resonanz von Kuhns Werk „The Structure of Scientific Revolutions“ gelang diese – primär auf die Naturwissenschaften bezogene – Thematik rasch in den Fokus sozialwissenschaftlicher Disziplinen, wobei die Organisationstheorie (resp. -soziologie) keine Ausnahme bildete (vgl. Hassard, 1993, S. 1). Abstrakt gesprochen bezeichnet Inkommensurabilität die Situation, in der für zwei Objekte im Hinblick auf eine zu vergleichende Dimension kein gemeinsames Mass für den Vergleich existiert (vgl. Kuhn, 1996, S. 103; Hoyningen-Huene, 1989, S. 203 f.).

Inkommensurabilität im Bereich der Organisationstheorie liegt – in einer ersten Annäherung – dann vor, wenn Methoden, Forschungsergebnisse und/oder Begriffe zwischen verschiedenen organisationstheoretischen Perspektiven in einem relevanten Sinne nicht miteinander vergleichbar sind (vgl. Scherer und Dowling, 1995, S. 202). Die Implikation derartiger Inkommensurabilitätsverhältnisse für die heterogene Landschaft der Organisationstheorie liegt auf der Hand: Die Perspektivenvielfalt führt nicht notwendigerweise zu einem vertieften und integrativen Verständnis organisationaler Phänomene, sondern resultiert stattdessen – wie die Kritiker betonen – in theoretischen und empirischen Integrationsproblemen unterschiedlicher Forschungsbemühungen (vgl. Tadajewski, 2009, S. 476; Burrell, 1996, S. 648). Die Frage, ob zwischen organisationstheoretischen Ansätzen Inkommensurabilitätsverhältnisse vorliegen, hat somit einen Einfluss auf die Bewertung der pluralen Theorielandschaft. Nimmt man eine reflexive Perspektive auf die Untersuchungen zu der Inkommensurabilitätsthematik in der Organisationstheorie ein, dann lassen sich die bisherigen Forschungsbemühungen wie folgt charakterisieren:

- Inhalt und Kritik der These inkommensurabler Perspektiven werden primär exegetisch entfaltet. Im Zentrum steht dabei die (zeitlich punktuell angelegte) Untersuchung der Schriften von Kuhn und Feyerabend (exemplarisch Tadajewski, 2009, S. 472 f.; Hassard, 1993, S. 76 ff.).
- Die begriffstheoretischen, epistemologischen und ontologischen Grundlagen des Begriffs der Inkommensurabilität werden im Zuge der Auseinandersetzung mit Kuhn und Feyerabend benannt, deren Beziehung untereinander jedoch ausgeblendet (exemplarisch Weaver und Gioia, 1994, S. 568 ff.; Morgan, 1980, S. 605 f.).
- Den Untersuchungen liegt ein digitales Verständnis von Inkommensurabilität zu Grunde. Aussagenzusammenhänge werden mehrheitlich entweder als inkommensurabel oder kommensurabel aufgefasst (exemplarisch McKinley et al., 1999, S. 634). Sporadisch findet sich die Differenzierung in eine starke und schwache Inkommensurabilität (exemplarisch Kirsch et al., 2007,

[1] Das subjektivistische Bild von (Sozial-)Wissenschaft ist antipositivistisch und systemisch-explorativ angelegt. Dagegen geht das objektivistische Bild von einem deterministischen, kausallogisch entschlüsselbaren Gegenstandsbereich der Sozialwissenschaften aus (vgl. Burrell und Morgan, 1979, S. 3).

[2] Strukturierungsansätze mit vergleichbar umfassendem Gliederungsanspruch finden sich bei Sorge (1997, S. 15 ff.), Evered und Louis (1981, S. 387 ff.) sowie Astley und van de Ven (1983, S. 246 ff.).

[3] Einschlägig sind hier insbesondere Kuhn (1962) und Feyerabend (1962).

S. 101).

• Das primäre organisationstheoretische Interesse besteht in der Handhabung inkommensurabler Perspektiven. Im Bereich der Theoriebildung und der Entwicklung von Forschungsmethoden wird aktiv versucht, Ansätze zu entwickeln, die eine Brücke zwischen den unterschiedlichen Theorieansätzen schlagen (exemplarisch Scherer und Steinmann, 1999, S. 524 ff.; Hassard, 1993, S. 82 ff., 88 ff.).

Diese Forschungsbemühungen lassen eine Forschungslücke offen. Wie ersichtlich ist, wird die Thematik der Inkommensurabilität in der Organisationtheorie vital diskutiert. Zudem werden Probleme, die durch dieses Phänomen für das Feld der Organisationstheorie auftreten können, benannt, ernstgenommen und Lösungsvorschläge erarbeitet. Was jedoch ausgespart bleibt, ist eine systematische Untersuchung des Begriffs der Inkommensurabilität, der über die reine Kuhn und Feyerabend Rezeption hinausgeht. An dieser Stelle leistet die vorliegende Untersuchung einen Beitrag. Grundlegende Motivation ist die Überzeugung, dass erst durch eine adäquate Fassung des Begriffs der Inkommensurabilität auch deren Konsequenzen für die Organisationstheorie klar erkennbar werden.

Die so verortete Untersuchung orientiert sich an zwei Forschungsfragen:

> Wie sieht ein Bezugsrahmen für die Thematik der Inkommensurabilität aus, der sich für organisationstheoretische Überlegungen eignet?
> An welchen Stellen und in welcher Form zeigen sich innerhalb der evolutionären Organisationstheorie Probleme, die durch Inkommensurabilität verursacht werden?

Die Struktur der Fragen legt eine gestufte Beantwortung nahe. In einem ersten Schritt (Kapitel 1) erfolgt eine systematische Klärung des Inkommensurabilitätsbegriffs. Es wird ein analytischer Bezugsrahmen erzeugt, der eine spezifische Perspektive für die angestrebte Begriffserläuterung eröffnet. Im Fokus steht die Frage, welche begriffstheoretischen und epistemologischen Thesen sich finden lassen, die als hinreichende Bedingung für das Zustandekommen einer Inkommensurabilitätsrelation anzusehen sind. Damit werden konstitutive Kriterien eines graduellen Begriffs von Inkommensurabilität erkennbar, die es erlauben, für eine konkrete Organisationstheorie festzustellen, wie ausgeprägt darin die Inkommensurabilitätsproblematik ist. In einem zweiten Schritt der Analyse (Kapitel 2) werden die Implikationen des gewonnenen Inkommensurabilitätskonzeptes für die evolutionäre Organisationstheorie erarbeitet. Der Beleuchtung der Konsequenzen wird eine Begründung der Wahl genau dieser Organisationstheorie sowie eine Skizze des evolutionären Theorieprogramms vorangestellt. Damit ist die Voraussetzung geschaffen, die Implikationen auf inter- und intratheoretischer Ebene deutlich zu machen.

Dieses wissenschaftstheoretisch angelegte Forschungsvorhaben macht Bemerkungen zur verwendeten Methodik notwendig. Um den Bezugsrahmen für die Erfassung der Inkommensurabilitätsproblematik aufzufalten, wird auf eine analytisch-philosophische Methode zurückgegriffen. Es wird die Methode der Begriffsklärung angewandt, um ein spezifisches Verständnis von Inkommensurabilität zu gewinnen: Inkommensurabilität bezeichnet eine Gruppe von Problemen, die in Situationen von Theorievergleichen auftreten können. Diese Begriffsklärung verläuft dementsprechend rein systematisch und damit losgelöst von spezifischen Texten und Autoren. Anleitend für diese Begriffsklärung ist die Frage, wie hinreichende Bedingungen aussehen müssen, damit sie für die Inkommensurabilitätsrelation konstitutiven Charakter haben. Der zweite Teil der Argumentation besteht in einem Transfer dieser Einsichten auf eine spezifische Organisationstheorie. Im Zuge dieses Transfers soll erstens geklärt werden, ob für die evolutionäre Organisationstheorie die Thematik der Inkommensurabilität überhaupt eine ausgeprägte Form annimmt. Zweitens wird es darum gehen, die Probleme, die mit dem Bestehen einer Inkommensurabilitätsrelation verbunden sind, konkret und möglichst detailliert für die Theorieanlage der evolutionären Organisationstheorie zu beschreiben. Zusammenfassend visualisiert Abbildung 1 die Struktur und das Forschungsvorhaben der vorliegenden Untersuchung.

Als Vorbemerkung für den Leser ist Folgendes vorauszuschicken. Die Thematik der Inkommensurabilität ist auf einer abstrakten, rein wissenschaftstheoretischen Ebene angesiedelt. Durch die Wahl der evolutionären Organisationstheorie als Transferobjekt wird der Abstraktionsgrad nur geringfügig verringert. Aus diesem Grund kommt insbesondere in Kapitel 2 der begrifflichen Klarheit und dem stringenten Aufbau der Untersuchung grösste Aufmerksamkeit zu.

2. Systematische Betrachtung der Inkommensurabilitätsrelation

Gegenstand dieses Kapitels ist eine systematische Diskussion der Inkommensurabilitätsthematik. Der Zusatz „systematisch“ deutet an, dass nicht die exegetische Auseinandersetzung mit einzelnen Schriften von Kuhn und Feyerabend im Zentrum steht. Stattdessen wird eine Herangehensweise verfolgt, die bewusst eine Vielzahl von Texten zur Inkommensurabilität einbezieht und im Sinne einer übergreifenden, vergleichenden Betrachtung den Begriff der Inkommensurabilität erläutert. Diese Begriffsklärung erfolgt entlang des folgenden Fragenkomplexes: Welche Bedingungen müssen erfüllt sein, so dass man überhaupt von inkommensurablen Theorien oder Problemen sprechen kann? Wie hängen diese Bedingungen miteinander zusammen?

Um diese Herangehensweise umzusetzen, ist das Kapitel wie folgt gegliedert. In Abschnitt 2.1 wird als Einstieg der Begriff der Inkommensurabilität in Kuhns Werk „The Structure of Scientific Revolutions“[4] verortet, um vier Aspekte des Begriffs herauszuarbeiten. Diese erlauben es Inkommensurabilität als zweigliedrige Relation zwischen Aussagensystemen

[4] Nachfolgend SSR abgekürzt.

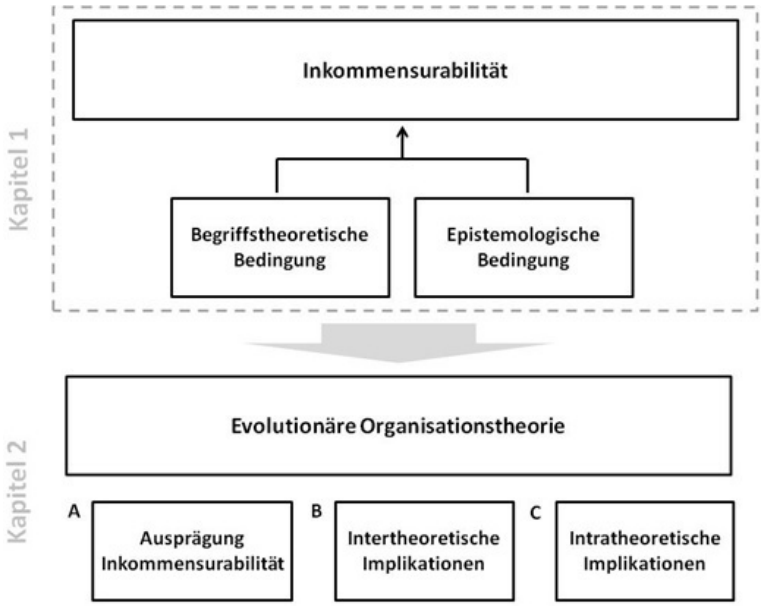

Abbildung 1: Struktur und Forschungsvorhaben der Untersuchung (Quelle: eigene Darstellung)

zu charakterisieren, mit der spezifische Probleme verbunden sind. In Abschnitt 2.2 wird ein metatheoretischer Bezugsrahmen für das in 2.1 spezifizierte Verständnis von Inkommensurabilität aufgespannt. Durch diesen Bezugsrahmen lässt sich ein Paket an begriffstheoretischen und epistemologischen Bedingungen für das Zustandekommen von Inkommensurabilität identifizieren. Ein Zwischenfazit verdichtet die gewonnenen Einsichten.

2.1. Ausgangs- und Ankerpunkt: Aspekte des Kuhnschen Begriffs von Inkommensurabilität

Wie einleitend bemerkt, erlebte SSR eine weitreichende Rezeption. Die Radikalität mit der Kuhn mit dem bisher vorherrschenden, kumulativen Bild von wissenschaftlicher Entwicklung brach, stiess neben positiven Rückmeldungen in einer Vielzahl wissenschaftlicher Disziplinen auch auf heftigen Widerstand in der Wissenschaftsphilosophie (vgl. Hoyningen-Huene, 1989, S. 7). Zugleich zu dem in SSR noch vagen Begriff des Paradigmas, war es insbesondere die Thematik der Inkommensurabilität, die den Kristallisationspunkt für die Kritiken bildete (vgl. ebd., S. 202). Die Redeweise von inkommensurablen Paradigmen und Theorien verleitete Kommentatoren dazu, Kuhn in Bezug auf den Theoriewahlprozess als Irrationalisten (exemplarisch Lakatos, 1970, S. 178) und Relativisten (exemplarisch Scheffler, 1967, S. 81) zu brandmarken.

Angesichts der Rezeptionsfülle und dem beschränkten Umfang der vorliegenden Untersuchung ist es sinnvoll, den Kuhnschen Begriff von Inkommensurabilität in kondensierter Form zu präsentieren.

Beginnen wir die Klärung des Begriffs mit zwei Textpassagen, die sich in SSR finden lassen:

> „The normal-scientific tradition that emerges from a scientific revolution is not only incompatible but often actually incommensurable with that which has gone before“ (Kuhn, 1996, S. 103, Hervorhebung d. d. V.)

Kuhn weiter:

> “To the extent (. . .) that two scientific schools disagree about what is a problem and what a solution, they will inevitably talk through each other when debating the relative merits of their respective paradigms” (ebd., S. 109, Hervorhebung d. d. V.)

Diesen beiden Textausschnitten kann eine erste Charakterisierung des Begriffs der Inkommensurabilität entnommen werden. Inkommensurabilität ist ein Phänomen, welches dem Themenkomplex der Theoriewahl zuzuordnen ist. Dabei generiert das Vorliegen von Inkommensurabilität offenbar Kommunikationsprobleme zwischen den Lagern, die an der Theoriewahl beteiligt sind. Ausgehend von dieser ersten Näherung werden nachfolgend vier Aspekte des Inkommensurabilitätsbegriffs bei Kuhn herausgearbeitet, die in der Summe zu einer hinreichend genauen Klärung des Begriffs für die weiteren Untersuchungsschritte führen.[5]

Erstens handelt es sich bei Inkommensurabilität um einen relationalen Begriff (vgl. Sankey, 1994, S. 19; Sankey und Hoyningen-Huene, 2001, S. viii). Das Vorliegen von Inkommensurabilität kann dementsprechend nur zwischen Objekten und nicht in Bezug auf ein einzelnes Objekt behauptet werden (vgl. Quine, 1969, S. 172 f.; Lorenz, 2004, S.

[5] Die vier Aspekte finden sich auch bei Feyerabend (1958, 1962, 1965, 1981, 1993). Aufgrund des beschränkten Umfangs der Untersuchung wird auf diese Textgrundlage verzichtet. Für eine ausführliche Darstellung von Feyerabends Inkommensurabilitätsbegriff siehe Oberheim und Hoyningen-Huene (2009, S. 10 ff.)

553). Inkommensurabilität funktioniert damit analog zum umgangssprachlichen Begriffspaar links/rechts. Die Aussage „Das Haus steht links“ ist kein sinnvoller deutscher Satz; die Behauptung „Anna steht rechts von Bernd“ dagegen schon. Wie die Begriffe links/rechts ist auch Inkommensurabilität ein zweigliedriger Relationsbegriff (vgl. Hoyningen-Huene, 1989, S. 203), d. h. Inkommensurabilität wird zwischen zwei Objekten konstatiert.[6] Spricht man demzufolge davon, dass eine Gruppe von Objekten (bspw. die Theorien A, B und C) inkommensurabel ist, so drückt man damit aus, dass jeweils die einzelnen Objektverhältnisse (im Beispiel A-B, A-C und B-C) inkommensurabel sind.

Zweitens lassen sich die Relata der Inkommensurabilitätsrelation als Aussagensysteme begreifen. Um diesen Punkt zu erläutern, ist es wiederum sinnvoll auf Kuhns initiale Darstellung von Inkommensurabilität zurückzugreifen. Er führt in SSR acht Typen von Objekte an, zwischen denen eine Inkommensurabilitätsrelation vorliegen kann (vgl. Preston, 2008, S. 89): Weltbilder, prä- und post-revolutionäre normalwissenschaftliche Traditionen, Erscheinungswelten, methodologische Standards, Paradigmen, Problemlösungen, Perspektiven und Theorien.[7] Diese Fülle von Relata ist für unsere Zwecke strukturierbar, wenn man die acht Typen jeweils als Aussagensysteme begreift. Diese Einschränkung, die mit Blick auf die weitere Argumentationslinie vollzogen wird, lässt sich wie folgt begründen. Betrachtet man die Entwicklung des Inkommensurabilitätsbegriffs in den Schriften Kuhns, so ist eine zunehmende Kopplung des Konzeptes der Inkommensurabilität an sprachliche Phänomene zu beobachten (vgl. Sankey, 1994, S. 22 ff.; Carrier, 2001, S. 65). In der Diskussion der Inkommensurabilität in SSR spielen visuelle Aspekte noch eine zentrale Rolle. Kuhn bemüht insbesondere die Gestaltsprung-Metaphorik[8], um aufzuzeigen, wie radikal verschieden inkommensurable Paradigmen, resp. Theorien sind (vgl. Kuhn, 1996, S. 112 ff.; Bird, 2000, S. 108 ff.). In späteren Schriften behandelt Kuhn Inkommensurabilität ausschliesslich im Rahmen von Übersetzungsproblemen zwischen Theorien (vgl. exemplarisch Kuhn, 1977b, S. xxii f.; Kuhn, 1982, S. 669, 673). Die Charakterisierung der Relata als Aussagensysteme trägt dementsprechend zum einen der historischen Entwicklungslinie der Kuhnschen Theorie Rechnung. Zum anderen lassen sich die acht Typen von Relata – insbesondere auch Weltbilder, Perspektiven und Traditionen – sprachlich beschreiben und damit mittels Aussagensystemen adäquat modellieren.

Drittens sind mit dem Bestehen einer Inkommensurabilitätsrelation zwischen zwei Aussagensystemen Probleme verbunden. Kuhn thematisiert in SSR drei Typen von Problemen (vgl. Hoyningen-Huene, 1989, S. 203). Zwischen inkommensurablen Theorien, resp. Paradigmen besteht zunächst ein methodologisches Problem. Es liegt keine Einigkeit darüber vor, was interessante und damit verfolgenswerte wissenschaftliche Probleme sind und wie Problemlösungen auszusehen haben (vgl. Kuhn, 1996, S. 103, 108). Des Weiteren ist ein begriffliches Problem zu konstatieren. Zwei inkommensurable Aussagensysteme können zwar Aussagen enthalten, in denen die gleichen Begriffe vorkommen. Dies ist jedoch noch nicht hinreichend dafür, dass sich die Aussagen ohne Probleme aufeinander beziehen lassen. Gemäss Kuhn muss die Bedeutung zweier identischer Begriffe innerhalb zweier inkommensurabler Aussagensysteme nicht identisch sein (vgl. ebd., S. 102). Zuletzt attestiert Kuhn, dass Vertreter inkommensurabler Paradigmen, resp. Theorien ihre Tätigkeit in unterschiedlichen Erlebniswelten ausführen (vgl. ebd., S. 111). Tritt man einen Schritt zurück und betrachtet das von Kuhn diskutierte Spektrum an Problemen, dann ist dieses wie folgt strukturierbar. Eine erste Klasse von Problemen kann mit dem Titel „Probleme der Verständigung“ überschrieben werden. Darunter fallen Übersetzungs- und Artikulationsprobleme. Ersterer Problemtyp umfasst die Frage, wie sich Aussagen der einen Theorie in eine Inkommensurable übersetzen lassen. Hinter dem zweiten Problemtyp steht die Frage, wie man Aussagen der beiden inkommensurablen Theorien sinnvoll aufeinander beziehen kann. Eine zweite Klasse von Problemen ist unter dem Titel „Probleme der epistemischen Rechtfertigung“ subsumierbar. Neben Problemen der Forschungsmethodik – was sind valable Forschungsmethoden, welche Probleme sind zu behandeln? – gibt es auch Probleme der Referenz zwischen zwei inkommensurablen Theorien. Diese lassen sich durch die Frage ausdrücken, ob zwischen inkommensurablen Theorien überhaupt Einigkeit über den zu behandelnden Gegenstand besteht.[9]

Viertens finden sich in den Schriften Kuhns sporadisch Hinweise darauf, dass die These einer Inkommensurabilitätsrelation an spezifische begriffstheoretische und epistemologische Überzeugungen geknüpft ist (siehe exemplarisch Kuhn, 1996, S. 96, 101 f., 113, 146; Kuhn, 1977a, S. 338; Kuhn, 1982, S. 671). Wie diese Überzeugungen auszubuchstabieren sind, ist Gegenstand von Abschnitt 2.2. An diesem Punkt genügt es festzuhalten, dass der Inkommensurabilitätsbegriff bei Kuhn aus begriffstheoretischen und epistemologischen Grundlagen entwickelt wird. Abbildung 2 visualisiert den bisher skizzierten Inkommensurabilitätsbegriff.

Fassen wir die gewonnenen Einsichten zusammen. Inkommensurabilität bezeichnet eine zweigliedrige Relation. Besteht diese Relation zwischen zwei Aussagensystemen,

[6] Diesbezüglich anderer Ansicht sind Scherer und Dowling (1995, S. 202). Nach unserer Ansicht vermischen Scherer und Dowling durch ihre multidimensionale Auffassung von Inkommensurabilität den relationalen Charakter des Inkommensurabilitätsbegriffs mit Problemen, die mit dem Bestehen von Inkommensurabilität verbunden sind. Siehe hierzu die Ausführungen zum dritten Aspekt.

[7] An dieser Stelle wird auf eine Charakterisierung der Begriffe Paradigma, Normalwissenschaft und wissenschaftliche Revolution verzichtet. Siehe hierfür Hoyningen-Huene, 1989, S. 133 ff., 165 ff., 193 ff.

[8] Eine wissenschaftliche Gemeinschaft, die eine wissenschaftliche Revolution durchlebt (so bspw. die Chemiker des frühen 18. Jahrhunderts), erlebt gemäss Kuhn einen Wechsel in ihrer visuellen Wahrnehmung, der analog zum plötzlichen Switch auf dem berühmten Bild der alten und jungen Dame verläuft (vgl. Hoyningen-Huene, 1989, S. 49).

[9] Diese Problemtypologie ist sehr verkürzt dargestellt und abstrakt angelegt. Der in Kapitel 2 vollzogene Transfer der Inkommensurabilitätsproblematik auf die evolutionäre Organisationstheorie bietet die Möglichkeit, die angesprochenen Probleme weiter auszuführen und zu exemplifizieren.

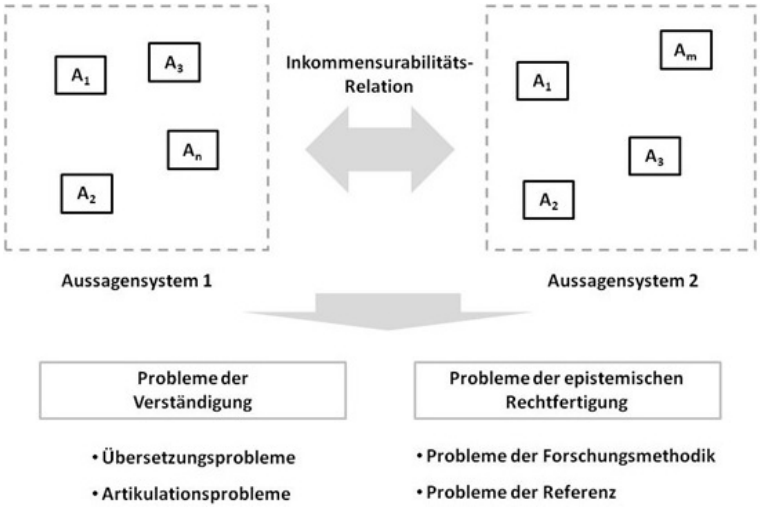

Abbildung 2: Inkommensurabilitätsbegriff (Quelle: eigene Darstellung)

dann treten Probleme der Verständigung sowie Probleme der epistemischen Rechtfertigung auf. Dieses Verständnis von Inkommensurabilität bildet den Ausgangs- und Ankerpunkt für die Überlegungen in Abschnitt 2.2.

2.2. Metatheoretischer Bezugsrahmen zur Inkommensurabilitätsrelation

Der metatheoretische Bezugsrahmen zur Inkommensurabilitätsrelation thematisiert die Frage, welche Bedingungen erfüllt sein müssen, damit eine Inkommensurabilitätsrelation vorliegen kann. Im Zuge der Argumentation wird eine begriffstheoretische und eine epistemologische Bedingung identifiziert, die zusammen hinreichend sind, um eine Inkommensurabilitätsrelation zu ermöglichen. Die Ausführungen sind wie folgt gegliedert. In Abschnitt 2.2.1 werden der Status und die Struktur des Bezugsrahmens diskutiert. Die Abschnitte 2.2.2 und 2.2.3 sind der schrittweisen Herleitung der begriffstheoretischen und epistemologischen Bedingungen gewidmet. In Abschnitt 2.2.4 werden die gewonnenen Einsichten verdichtet und der Bezugsrahmen so präsentiert, dass er für die evolutionäre Organisationstheorie fruchtbar gemacht werden kann.

2.2.1. Vorbemerkungen – Status und Struktur des Bezugsrahmens

In Abschnitt 2.1 wurde ein relationaler Begriff von Inkommensurabilität gewonnen. Mit diesem Begriff kann nun auf zwei grundsätzlich verschiedene Arten verfahren werden. Man kann sich die Frage stellen, ob tatsächlich Aussagensysteme gefunden werden können, zwischen denen eine Inkommensurabilitätsrelation vorliegt. Verfolgt man einen derartigen Ansatz, dann ist bspw. zu untersuchen, ob sich im paarweisen Vergleich von Organisationstheorien Situationen finden lassen, in denen zwischen den Vertretern der Theorien Uneinigkeiten hinsichtlich der Forschungsmethodik sowie des Gegenstandbereichs bestehen und/oder inwieweit Verständigungsschwierigkeiten vorliegen. Eine solche Herangehensweise kann als objekttheoretischer Zugang zur Inkommensurabilitätsrelation bezeichnet werden. Im Fokus dieses Zugangs steht abstrakt gesprochen die Objektklasse von Aussagensystemen und deren Beziehungen untereinander. Anleitend für die Untersuchung dieser Objektklasse ist ein feststehender Begriff von Inkommensurabilität der angewandt und nicht weiter reflektiert wird.[10]

In der vorliegenden Untersuchung wird demgegenüber eine metatheoretische Perspektive auf die Inkommensurabilitätsrelation eingenommen. Dabei steht die Relation im Fokus der Überlegungen. Anstatt diese auf die Objektklasse der Aussagensysteme anzuwenden, wird untersucht, welche begriffstheoretischen und epistemologischen Positionen man einnehmen muss, um überhaupt von einer Inkommensurabilitätsrelation sprechen zu können. Es ist dementsprechend ein Zugang, welcher die theoretischen Grundlagen der Inkommensurabilitätsrelation thematisiert.

Der zu entwickelnde Bezugsrahmen ist ein Paket aus zwei Bedingungen. Eine begriffstheoretische und eine epistemologische Bedingung bilden zusammen ein hinreichendes Paket an Kriterien, das es erlaubt für ein beliebiges Theoriekonstrukt festzustellen, inwiefern dieses mit der Inkommensurabilitätsthematik konfrontiert ist. Der Bezugsrahmen ermöglicht es somit aufgrund der begriffstheoretischen und epistemologischen Zugeständnissen, die ein Theoriekonstrukt eingeht, die Ausprägung der Inkommensurabilitätsthematik für diese Theorieanlage zu bestimmen.

Vor dem Hintergrund dieser Bemerkungen lässt sich ein Aspekt der vorliegenden Untersuchung entschärfen, der auf den ersten Blick irritiert. Wie in der Einleitung ausgeführt, werden in Kapitel 3 die Implikationen der Inkommensurabilität auf eine spezifische Organisationstheorie herausgearbeitet. Die evolutionäre Organisationstheorie wird nicht in Beziehung zu einer zweiten Organisationstheorie gesetzt, um

[10] Beispiele für einen derartigen objekttheoretischen Zugang in der Organisationstheorie sind Burrell (1997); Donaldson (1985); Jackson und Carter (1991); Weaver und Gioia (1994) sowie McKinley und Mone (1998).

eine Inkommensurabilitätsrelation zu identifizieren. Stattdessen wird ausgehend von den begriffstheoretischen und epistemologischen Zugeständnissen, welche die evolutionäre Organisationstheorie eingeht, untersucht, wie ausgeprägt die Inkommensurabilitätsthematik für dieses Theoriekonstrukt ist und an welchen Stellen in der Theorieanlage sich die damit verbundenen Probleme artikulieren. Berücksichtigt man die Unterscheidung zwischen einem objekt- und einem metatheoretischen Zugang, verschwindet das irritierende Moment in der Anlage der Untersuchung.[11] Zusammenfassend visualisiert Abbildung 3 den Status und die Struktur des Bezugsrahmens.

Wie in Abbildung 3 erkennbar ist, bilden die zwei Bedingungen ein Kriterienpaket, welches sowohl notwendig als auch hinreichend für das Bestehen einer Inkommensurabilitätsrelation ist. Aus diesem Grund gliedert sich die nachfolgende Argumentation in zwei Schritte. In Abschnitt 2.2.2 wird aufgezeigt, dass die Bedingung 1 und 2 notwendige Bedingungen darstellen (d. h. es wird die Aussagerichtung '→' diskutiert). In Abschnitt 1.2.3 findet sich die Erläuterung dafür, dass die Bedingungen 1 und 2 zusammen hinreichend sind (d. h. die Aussagerichtung '←' wird dargelegt). In Kombination resultiert aus diesen beiden Argumentationsschritten das angestrebte Kriterienpaket.

2.2.2. Schritt 1 – Notwendige Bedingungen

Das Ziel des ersten Schritts in der Konstruktion des Bezugsrahmens besteht darin, die Inkommensurabilitätsrelation in zwei notwendige Bedingungen zu analysieren. Dieses Vorhaben lässt sich durch die folgenden zwei Thesen weiter präzisieren:

> *These 1:* Die Möglichkeit einer Inkommensurabilitätsrelation zwischen zwei Aussagensystemen impliziert eine nicht-augustinische Begriffstheorie.
> *These 2:* Die Möglichkeit einer Inkommensurabilitätsrelation zwischen zwei Aussagensystemen impliziert eine nicht-positivistische Epistemologie.

Diese beiden Thesen enthalten die gesuchten notwendigen Bedingungen, die an die Inkommensurabilitätsrelation gekoppelt sind:

> *Notwendige Bedingung (1):* Nicht-augustinische Begriffstheorie
> *Notwendige Bedingung (2):* Nicht-positivistische Epistemologie

Die beiden Bedingungen sind in dieser kondensierten Form nicht verständlich. Um zu verstehen, was damit gemeint ist, müssen drei Dinge geklärt werden. Erstens ist zu präzisieren, was der Zusatz „nicht" ausdrücken soll. Zweitens stellt sich die Frage, was die zentralen Eigenschaften einer augustinischen Begriffstheorie sind. Die analoge Frage kann drittens auch im Hinblick auf die positivistische Epistemologie gestellt werden. Nachfolgend werden diese drei Punkte geklärt und anschliessend zwei Argumente präsentiert, die aufzeigen, dass die beiden Bedingungen tatsächlich notwendig sind.

Der Zusatz *„nicht"* drückt aus, dass nur diejenigen Begriffstheorien und Epistemologien die notwendigen Bedingungen erfüllen, welche die wesentlichen Punkte der augustinischen Begriffstheorie und der positivistischen Epistemologie nicht teilen. Konkret ist damit folgende Forderung verbunden: Gegeben man behauptet die Möglichkeit einer Inkommensurabilitätsrelation zwischen zwei Aussagensystemen, dann muss man gleichzeitig auch eine Begriffstheorie und eine Epistemologie vertreten, welche die wesentlichen Merkmale mit der augustinischen Begriffstheorie, resp. der positivistischen Epistemologie nicht teilen. Um diese Forderung weiter diskutieren zu können, sind die beiden philosophischen Positionen zu charakterisieren.

Die *augustinische Begriffstheorie* erhielt ihre philosophische Relevanz durch die Erwähnung im ersten Paragraphen von Wittgensteins Philosophischen Untersuchungen (Wittgenstein, 1999, S. 237). Wittgenstein nutzt die augustinische Begriffstheorie als Folie, vor der er seine Gebrauchstheorie der Bedeutung entwickelt (vgl. Glock, 1996, S. 41 ff.).[12] Zu den Charakteristika dieser Begriffstheorie zählen die Folgenden: Ein Begriff (hier als sprachlicher Ausdruck aufgefasst) erhält seine Bedeutung durch das Objekt, welches er bezeichnet (vgl. ebd., S. 41). Der Begriff wird dabei durch eine hinweisende Definition mit dem bedeutungsstiftenden, materiellen Objekt verbunden (vgl. Schulte, 2001, S. 171; Backer und Hacker, 2005, S. 4). Als Beispiel kann das Wort „Tisch" dienen. Wir verstehen den Begriff „Tisch" – können ihm Bedeutung zuschreiben – nur, weil wir den Begriff durch eine hinweisende Definition mit einem bestimmten Objekt (bspw. einem konkreten Tisch A) verknüpft haben. Sätze sind nach der augustinischen Begriffstheorie Aggregate von Begriffen und erhalten ihren Sinn durch die Bedeutung der Begriffe, aus denen sie zusammengesetzt sind (vgl. Glock, 1996, S. 41). Sprache steht dementsprechend sowohl auf Ebene der Begriffe, wie auch auf Ebene der Sätze in einer ausschliesslich deskriptiven Beziehung zur Welt (vgl. ebd.).

Das zentrale Merkmal der augustinischen Begriffstheorie besteht darin, dass die Bedeutung von Begriffen durch Objekte etabliert wird, auf die sich die Begriffe beziehen. Diese Menge von Objekten ist die Grundbasis für die Verwendung von Sprache (vgl. ebd.). Nicht sprachliche Objekte stellen dementsprechend einen materiellen Bezugspunkt dar, der

[11] Der Leser kann bereits an dieser Stelle einwenden, dass nicht jede Organisationstheorie ihr begriffstheoretisches und epistemologisches Fundament offen artikuliert. Damit biete sich dem metatheoretischen Bezugsrahmen keine Einstiegsfläche. Gegen diesen Einwand spricht zum einen, dass in die Forschungsmethoden und Aussagen von Organisationstheorien implizit stets wissenschaftstheoretische Grundannahmen eingehen. Zum andern ist zuzugeben, dass sich dieser Ansatz speziell für Organisationstheorien eignet, die ihr eigenes Fundament thematisieren. Dass es sich bei der evolutionären Organisationstheorie um eine Theorie dieses Typs handelt, wird in Abschnitt 3.1 ausführlich dargelegt.

[12] In Abschnitt 3.2 findet sich eine Kurzcharakterisierung der wittgensteinschen Begriffstheorie.

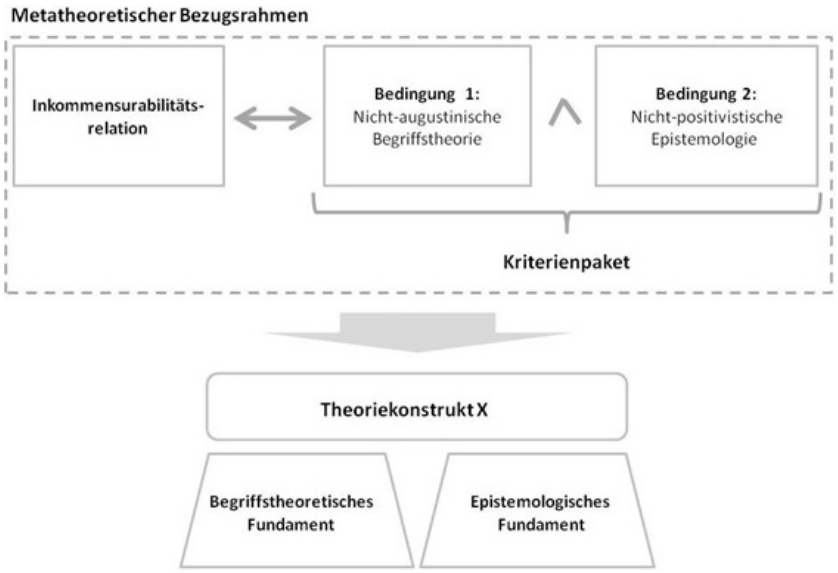

Abbildung 3: Status und Struktur des metatheoretischen Bezugsrahmens zur Inkommensurabilitätsrelation (Quelle: eigene Darstellung; Die zwei aussagenlogischen Zeichen "↔" und "∧" werden in Online-Appendix A erläutert.)

zwei zentrale Eigenschaften der Bedeutung von Begriffen sicherstellt. Die Bedeutung eines Begriffs ist erstens kontextunabhängig und damit stabil. Der Begriff „Tisch" hat nur eine Bedeutung, egal ob er während einer alltäglichen Konversation oder in einer detaillierten Beschreibung der Innenausstattung des Schlosses von Versailles im frühen 18. Jahrhundert verwendet wird. Dieser Bezugspunkt ist zweitens zugleich eine Ebene auf die intersubjektiv referiert werden kann. Liegen Verständigungsschwierigkeiten zwischen zwei Personen oder Gruppen von Personen vor, dann kann auf das Faktum verwiesen werden, dass es eine Ebene von Objekten gibt, die für die Bedeutung eines Begriffs relevant ist. Diese Ebene ist intersubjektiv zugänglich und somit idealer Ausgangspunkt für eine Konsensfindung zwischen den Gesprächsteilnehmern.

Wenden wir uns der Erläuterung der *positivistischen Epistemologie* zu. Der Ausdruck „positivistische Epistemologie" dient uns als Sammelbegriff für zentrale Merkmale der Tradition des Positivismus im frühen 20. Jahrhundert.[13] Mit dieser Tradition sind zwei erkenntnistheoretische Thesen verbunden. Zum einen vertreten die Positivisten eine spezifische Wahrnehmungstheorie (vgl. Macdonald, 2005, S. 6). Wahrnehmung liefert uns stets Wahrnehmungsinhalte, die als intersubjektiv gegeben angenommen werden können. Diese gegebenen Sinneseindrücke fungieren als Rohmaterial für alle weiteren mentalen Operationen, durch die komplexere Wahrnehmungen synthetisiert werden (vgl. Uebel, 2006, S. 21). Zum anderen betonen die Positivisten eine bestimmte Form der epistemischen Rechtfertigung (vgl. Hacking, 1996, S. 77). Alle Wissensansprüche, die empirischen Gehalt aufweisen (d. h. diejenigen Aussagen, die sich auf die erfahrbare Welt beziehen), müssen sich mit Bezug auf Aussagen über elementare Beobachtungen rechtfertigen lassen. Einzig diese reduktive Art der epistemischen Rechtfertigung ist für empirische Aussagen zulässig (vgl. ebd.) Ein Beispiel kann diesen Punkt verdeutlichen. Die Aussage „Peter arbeitet um 8.00 Uhr an seinem Arbeitsplatz bei der SBB" drückt einen Wissensanspruch über die empirisch zugängliche Welt aus. Derjenige Sprecher, der diese Aussage äussert und sie als Wissen taxiert, muss eine Rechtfertigung für seinen Wissensanspruch liefern können. Diese Rechtfertigung besteht gemäss dem vorab Gesagten in der Angabe einer Menge von Beobachtungsaussagen, welche die Richtigkeit der ursprünglichen Aussage belegen. Im vorliegenden Fall wären dies bspw. unter anderen die Aussagen „Objekt X sitzt" und „Objekt X befindet sich im SBB Gebäude".

Wiederum können wir nach dem zentralen Merkmal dieser epistemologischen Position fragen. Dieses besteht darin, dass sowohl in Bezug auf unseren Wahrnehmungsprozess wie auch in Bezug auf die epistemische Rechtfertigung unserer Wissensansprüche eine intersubjektiv zugängliche Ebene vorausgesetzt wird. Wahrnehmung – so die Positivisten – beginnt stets bei primitiven Wahrnehmungsinhalten. Existieren Diskrepanzen bezüglich Wahrnehmungen, so können diese – zumindest prinzipiell – durch eine Analyse derselben in die sie konstituierenden Wahrnehmungsinhalte aufgelöst werden. Analoges gilt für die epistemische Rechtfertigung. Einander widersprechende Wissensansprüche über einen bestimmten Gegenstand lassen sich durch den Rekurs auf die Ebene der Beobachtungsaussagen – zumindest prinzipiell – entscheiden.

Durch die Erläuterung der drei Komponenten „nicht", „augustinische Begriffstheorie" und „positivistische Epistemologie" sind die notwendigen Bedingungen (1) und (2) zugänglich geworden. Im nächsten Schritt geht es darum aufzuzeigen, dass es sich bei den beiden Bedingungen tat-

[13] Für eine detaillierte Diskussion der Tradition des Positivismus siehe Uebel (2006).

sächlich um notwendige Bedingungen für die Inkommensurabilitätsrelation handelt.

Beginnen wir mit der *notwendigen Bedingung (1);* der nicht-augustinischen Begriffstheorie. Damit diese Bedingung eine notwendige Bedingung für die Inkommensurabilitätsrelation darstellt, gilt es folgende These zu plausibilisieren:

> *These 1:* Die Möglichkeit einer Inkommensurabilitätsrelation zwischen zwei Aussagensystemen impliziert eine nicht-augustinische Begriffstheorie.

Diese These drückt dasselbe aus wie die Folgende:[14]

> *These 1‘:* Der augustinischen Begriffstheorie zufolge ist eine Inkommensurabilitätsrelation zwischen zwei Aussagensystemen nicht möglich.

Diese These 1‘ ist direkter einzusehen als die Ausgangsthese und kann wie folgt argumentativ gestützt werden. Als zentrales Merkmal der augustinischen Begriffstheorie haben wir eine intersubjektiv zugängliche Ebene von bedeutungsstiftenden Objekten identifiziert. Diese Ebene sorgt dafür, dass Begriffe über verschiedene Kontexte hinweg die gleiche Bedeutung besitzen. Darüber hinaus ist diese Ebene ein Bezugspunkt für die Feststellung der richtigen Bedeutung eines Begriffs, falls ein sprachliches Missverständnis zwischen zwei oder mehreren Sprechern vorliegt. Bei der Charakterisierung der Inkommensurabilitätsrelation aus Abschnitt 1.1 wurde deutlich, dass mit dem Inkommensurabilitätsbegriff spezifische Probleme verbunden sind. Unter dem Stichwort „Probleme der Verständigung“ wurden Übersetzungs- und Artikulationsprobleme diskutiert, die auftreten, sobald eine Inkommensurabilitätsrelation besteht. Setzt man nun eine augustinische Begriffstheorie voraus, können diese zwei Typen von Problemen nicht mehr auftreten. Übersetzungsprobleme zwischen zwei Sprechern oder Gruppen von Sprechern liegen nicht vor, da Begriffe über verschiedene Kontexte, resp. Aussagensysteme hinweg eine stabile Bedeutung besitzen. Artikulationsprobleme verschwinden ebenfalls, da eine intersubjektiv zugängliche Ebene vorhanden ist, auf die Bezug genommen werden kann, falls sich Probleme beim Vergleich von verschiedenen Aussagen ergeben. Setzt man dementsprechend eine augustinische Begriffstheorie voraus, dann können die für eine Inkommensurabilitätsrelation konstitutiven Probleme der Verständigung nicht auftreten. Damit ist die Möglichkeit einer Inkommensurabilitätsrelation ausgeschlossen.

Wenden wir uns der notwendigen Bedingung (2), der nicht-positivistischen Epistemologie, zu. Damit diese Bedingung eine notwendige Bedingung für eine Inkommensurabilitätsrelation darstellt, ist folgende These zu untermauern:

> *These 2:* Die Möglichkeit einer Inkommensurabilitätsrelation zwischen zwei Aussagensystemen impliziert eine nicht-positivistische Epistemologie.

Diese These besagt dasselbe, wie die Folgende:[15]

> *These 2‘:* Der positivistischen Epistemologie zufolge ist eine Inkommensurabilitätsrelation zwischen zwei Aussagensystemen nicht möglich.

Wiederum ist die These 2‘ leichter einzusehen als die Ausgangsthese und lässt sich durch die folgenden Überlegungen plausibilisieren. Als zentrales Merkmal der positivistischen Epistemologie wurde vorgängig eine intersubjektiv zugängliche Ebene von primitiven Wahrnehmungsinhalten identifiziert. Diese Ebene ist sowohl Ausgangspunkt für unsere Wahrnehmungen der Umwelt wie auch Basis für die Rechtfertigung unserer Wissensansprüche über die erfahrbare Wirklichkeit.

Neben den bereits besprochenen Problemen der Verständigung diskutierten wir in Abschnitt 2.1 Probleme der epistemischen Rechtfertigung, die mit der Inkommensurabilitätsrelation verbunden sind. Darunter fallen Probleme der Forschungsmethodik sowie Probleme der Referenz. Wiederum gilt: Sobald eine positivistische Epistemologie vorausgesetzt wird, können diese zwei Problemtypen nicht mehr auftreten. Die Frage, welche Forschungsmethoden angebracht sind, lässt sich mit Bezug auf die Ebene der prinzipiell intersubjektiv zugänglichen Wahrnehmungsinhalte sinnvoll diskutieren. Diskussionsbarrieren zwischen Befürwortern unterschiedlicher Forschungsmethoden bestehen nicht, wenn ausschliesslich eine, unumstrittene Datengrundlage vorliegt. Das Problem der Referenz verschwindet gleichermassen, da mit der Ebene der Wahrnehmungsinhalte ein intersubjektiv zugänglicher Gegenstandsbereich vorliegt, der den Forschungen zu Grunde gelegt werden kann. Setzt man dementsprechend eine positivistische Epistemologie voraus, verschwinden die für die Inkommensurabilitätsrelation konstitutiven Probleme der epistemischen Rechtfertigung. Somit ist auch in diesem Fall die Möglichkeit einer Inkommensurabilitätsrelation ausgeschlossen.

Zusammenfassend lässt sich festhalten, dass sich die Inkommensurabilitätsrelation in die zwei notwendigen Bedingungen „nicht-augustinische Begriffstheorie“ sowie „nicht-positivistische Epistemologie“ analysieren lässt. Diese abstrakte Einsicht wird nachfolgend anhand eines *ersten Fallbeispiels* verdeutlicht.

Kuhns Ansatz zur Beschreibung und Erklärung der wissenschaftlichen Entwicklung ist ein Theoriekonstrukt, in dem Inkommensurabilität eine tragende Rolle spielt (vgl. Hoyningen-Huene, 1989, S. 202; Brown, 2001, S. 126). Gemäss der vorangegangenen Analyse müssten Kuhns Überlegungen dementsprechend auf einer Begriffstheorie und einer Epistemologie basieren, welche die beiden notwendigen Bedingungen erfüllen.

[14] Dass sich die zwei Thesen entsprechen ist intuitiv nicht direkt erkennbar. In Online-Appendix B wird aufgezeigt, dass die Thesen 1 und 1‘ tatsächlich dasselbe ausdrücken.

[15] Wiederum ist intuitiv nicht direkt klar, dass sich die beiden Thesen entsprechen. In Online-Appendix B wird aufgezeigt, dass die Thesen 2 und 2‘ tatsächlich dasselbe ausdrücken.

Betrachtet man die Schriften Kuhns, dann finden sich Anhaltspunkte, die dies nahelegen. Die Überlegungen Kuhns basieren zum einen auf einer spezifischen Begriffstheorie. Sowohl in den frühen (exemplarisch Kuhn, 1996, S. 102) wie auch in den späten Schriften (exemplarisch Kuhn, 1982, S. 670 f.) findet sich eine Kontexttheorie der Bedeutung (vgl. Carrier, 2008, S. 46). Gemäss dieser erhalten Begriffe ihre Bedeutung nicht durch ein Objekt in der Umwelt, sondern durch den Aussagenzusammenhang, in dem sie vorkommen (vgl. Kuhn, 1996, S. 102, 146, 149; Kuhn, 1977a, S. 338; Kuhn, 1977b, S. xiv; Boyd, 2001, S. 5). Die Einbettung eines Begriffs in ein Netzwerk von Begriffen macht die legitimen inferentiellen Übergänge zwischen Begriffen deutlich und bestimmt damit die jeweilige Bedeutung der Begriffe (vgl. Carrier, 2001, S. 67; Hacking, 1996, S. 127). Diese Kontexttheorie gilt gemäss dem frühen Kuhn für alle theoretischen Begriffe. Darunter fallen im naturwissenschaftlichen Gebrauch Begriffe wie Masse, Bewegung und Elektron (vgl. Hacking, 1996, S. 101; Carrier, 2001, S. 67). In den späten Schriften ist der Kontext nur noch für eine Untermenge dieser theoretischen Begriffe bedeutungsstiftend (vgl. Hoyningen-Huene, 1989, S. 115, 159; Sankey, 1994, S. 28; Preston, 2008, S. 94; Chen, 1997, S. 257 ff.). Der zentrale Punkt für unsere Überlegung bleibt jedoch erhalten. Sowohl in den frühen wie auch späten Texten von Kuhn spielt Inkommensurabilität eine tragende Rolle. Gleichzeitig findet sich mit der Kontexttheorie der Bedeutung eine begriffstheoretische Position, welche eine nicht-augustinische Begriffstheorie darstellt. In dem Kuhn zulässt, dass Begriffe über unterschiedliche Kontexte hinweg keine stabile Bedeutung haben, verwirft er die zentrale Annahme der augustinischen Begriffstheorie.

Zum andern distanziert sich Kuhn von der positivistischen Strömung innerhalb der Wissenschaftstheorie (vgl. exemplarisch Kuhn, 1996, S. 96; Preston, 2008, S. 53; Bird, 2000, S. 9 ff.). So findet sich in SSR die These von der Theoriebeladenheit der Beobachtung (vgl. Kuhn, 1996, S. 111 f., 113, 120; Bird, 2000, S. 108 ff., 155). Kuhn erläutert in Anlehnung an Hanson (1965), dass unsere Wahrnehmung durch den begrifflichen Rahmen, welcher uns zur Verfügung steht, beeinflusst und strukturiert wird (vgl. Kuhn, 1996, S. 112 f.; Kuhn, 1977c, S. 308 f.; Irzik und Grünberg, 1998, S. 208 ff.). Verweise auf wahrnehmungspsychologische Experimente untermauern bei Kuhn diesen Punkt und machen deutlich, dass Wahrnehmung kein passiver Vorgang darstellt, sondern aufgrund der Rolle begrifflicher Kategorien immer auch subjektseitige Momente enthält (vgl. Kuhn, 1996, S. 120, 125; Kuhn, 1977d, S. 226; Preston, 2008, S. 64). Obwohl sich Kuhn in späteren Schriften von der Gestalt-Metaphorik für Wahrnehmungsprozesse verabschiedete (vgl. Hoyningen-Huene, 1989, S. 68), blieb der wesentliche Aspekt für unsere Argumentation erhalten. Die These von der Theoriebeladenheit der Beobachtung stellt eine nicht-positivistische Epistemologie dar. Indem Kuhn einen Einfluss der begrifflichen Ebene auf unsere Wahrnehmungseindrücke attestiert, verwirft er die zentrale Annahme der positivistischen Epistemologie: die Vorstellung, dass es einen intersubjektiv zugänglichen Bereich von primären Sinneseindrücken gibt, der frei ist von jeglicher Beeinflussung durch unsere Begriffe und als Rohmaterial für mentale Operationen dient.

Zusammenfassend können wir festhalten, dass in Kuhns Theorieanlage Inkommensurabilität eine tragende Rolle spielt. Dabei – und dies ist der zentrale Punkt – basieren seine Überlegungen auf einer begriffstheoretischen und epistemologischen Position, welche die beiden notwendigen Bedingungen erfüllen, die wir im Rahmen unserer Analyse gefunden haben.

2.2.3. Schritt 2 – Hinreichende Bedingungen

Das Ziel des zweiten Schritts in der Konstruktion des Bezugsrahmens besteht darin aufzuzeigen, dass die in 2.2.2 identifizierten notwendigen Bedingungen zusammengenommen das Bestehen der Inkommensurabilitätsrelation ermöglichen.

Um diesen argumentativen Schritt zu vollziehen, greifen wir auf eine Plausibilitätsüberlegung zurück. Die nachfolgende These legt nahe, dass die beiden notwendigen Bedingungen zusammengenommen hinreichend sind, um die Inkommensurabilitätsrelation und die damit verbundenen Probleme hervorzurufen:

> Je weiter man sich argumentativ von den beiden notwendigen Bedingungen „nicht-augustinische Begriffstheorie“ und „nicht-positivistische Epistemologie“ entfernt, desto ausgeprägter wird die Inkommensurabilitätsrelation zwischen zwei Aussagensystemen.

Die These ist in dieser kondensierten Form nicht verständlich. Aus diesem Grund wird sie nachfolgend präzisiert, bevor sie untermauert und durch ein zweites Fallbeispiel veranschaulicht wird.

Was bedeutet es, sich argumentativ von einer bestimmten Position zu entfernen? Im Kontext des hier diskutierten metatheoretischen Bezugsrahmens ist damit die Tatsache angesprochen, dass man die zentralen Merkmale einer bestimmten Position (bspw. diejenigen der augustinischen Begriffstheorie) nicht vollständig übernehmen muss, sondern partiell zurückweisen kann. Sich von einer Position argumentativ zu entfernen meint somit ein graduell gestufter Prozess, bei dem grössere oder kleinere Teile der entscheidenden Merkmale einer spezifischen Position übernommen werden. In unserem Fall lässt sich ein zentrales Merkmal in Bezug auf die augustinische Begriffstheorie und positivistische Epistemologie finden. Beide Theoriekonstrukte postulieren eine Ebene, die intersubjektiv zugänglich ist und auf die referiert werden kann, falls Probleme der Verständigung oder der epistemischen Rechtfertigung auftreten. Dieses zentrale Merkmal kann entweder vollständig oder in graduell abgeschwächter Form als Paket in das wissenschaftstheoretische Fundament der eigenen Theoriebemühungen eingehen.

In einem weiteren Schritt ist zu klären, inwiefern davon gesprochen werden kann, dass die Ausprägung einer Inkommensurabilitätsrelation zunimmt. Diese Redeweise impliziert ein graduelles Verständnis von Inkommensurabilität. Eine zu-

nehmende Ausprägung der Inkommensurabilitätsrelation bedeutet, dass die damit verbundenen Probleme in einem stärkeren Masse auftreten. Die in Abschnitt 2.1 thematisierten Probleme der Verständigung und der epistemischen Rechtfertigung werden also deutlicher sichtbar.

Wie lässt sich dieses graduelle Verständnis von Inkommensurabilität untermauern? Wir haben in Abschnitt 2.2.2 gesehen, dass eine augustinische Begriffstheorie sowie eine positivistische Epistemologie eine intersubjektiv zugängliche Ebene etablieren, welche das Auftreten einer Inkommensurabilitätsrelation verhindert. Entfernt man sich argumentativ entlang der begriffstheoretischen und epistemologischen Dimension, dann lässt man Aspekte dieser intersubjektiven Ebene fallen. Ohne festen Bezugspunkt für Bedeutungen und Wahrnehmungsinhalte können in der Folge die Probleme auftreten, die konstitutiv für die Inkommensurabilitätsrelation sind. Je weniger man an der Idee festhält, dass Bedeutungen über verschiedene Kontexte hinweg konstant bleiben, desto mehr Bereiche von Aussagensystemen sind durch Bedeutungsunterschiede gekennzeichnet und desto mehr Probleme der Verständigung treten auf. Analog gilt: Je weniger man an der These festhält, dass Wahrnehmungsinhalte intersubjektiv geteilt werden, desto problematischer wird die Vorstellung von einem intersubjektiv zugänglichen Gegenstandsbereich, der den gemeinsamen Forschungsbemühungen zu Grunde gelegt wird. Damit nehmen die Probleme der epistemischen Rechtfertigung zu.

In Abbildung 4 sind diese Einsichten visualisiert. Bewegt man sich entlang der begriffstheoretischen und epistemologischen Dimension in Pfeilrichtung weg vom Ausgangspunkt, dann nimmt die Ausprägung der Inkommensurabilitätsrelation zu. Dies wird durch die Notation $I = 0, I = 1, \ldots, I = n$ zum Ausdruck gebracht, wobei $I = 0$ bspw. bedeutet, dass keine Inkommensurabilitätsrelation auftreten kann. Die gestrichelten Linien stellen Isolinien dar. Bewegt man sich auf einer Isolinie, dann bleibt die Ausprägung der Inkommensurabilitätsrelation konstant. Die Hyperbelform der Isolinien deutet an, dass es Substitutionsmöglichkeiten zwischen der begriffstheoretischen und der epistemologischen Position gibt. Konkret sind die Isolinien ein Vorschlag, wie man das Verhältnis zwischen Begriffstheorie und Epistemologie im Hinblick auf die Inkommensurabilitätsrelation fassen kann. Der Aspekt der Substitution wurde nicht diskutiert und ist für den weiteren Verlauf der Untersuchung nicht von Relevanz. Zentral ist dagegen die Einsicht, dass die Inkommensurabilitätsproblematik zunimmt, wenn man sich entlang der Pfeilrichtung weg vom Ausgangspunkt bewegt. Diese Einsichten, die wiederum auf einer abstrakten Ebene angelegt sind, werden nachfolgend anhand eines *zweiten Fallbeispiels* veranschaulicht.

Feyerabend gilt neben Kuhn als prägend für den Begriff der Inkommensurabilität (vgl. Hoyningen-Huene, 1990, S. 482). Für unsere Zwecke ist die Frage interessant, unter welchen spezifischen Voraussetzungen Kuhn und Feyerabend Inkommensurabilität thematisieren und ob sich daraus ablesen lässt, für wen Inkommensurabilität und die damit verbundenen Probleme eine ausgeprägtere Form annehmen.

Kuhns begriffstheoretische und epistemologische Position wurde im ersten Fallbeispiel ausführlich dargelegt. Wir haben festgestellt, dass Kuhn für abstrakte theoretische Begriffe eine Kontexttheorie der Bedeutung voraussetzt. Zudem postuliert Kuhn eine Theoriebeladenheit der Wahrnehmungsprozesse. Kontrastiert man diese beiden Voraussetzungen mit den begriffstheoretischen und epistemologischen Annahmen, die sich bei Feyerabend finden lassen, dann kann man zwei Punkte festhalten. Erstens vertritt Feyerabend in Bezug auf die epistemologische Dimension eine gleich starke These wie Kuhn (vgl. Oberheim und Hoyningen-Huene, 1997, S. 450). Auch Feyerabend teilt die neo-kantianische Intuition, dass Begriffe als Ordnungsschemata einen Einfluss auf die Inhalte unserer Wahrnehmung ausüben (vgl. Feyerabend, 1958, S. 148; Feyerabend, 1981, S. 44 f.; Oberheim und Hoyningen-Huene, 2009, S. 16 f., 19). Zweitens findet sich bei Feyerabend eine extremere begriffstheoretische Position als bei Kuhn. Die Kontexttheorie der Bedeutung gilt, so Feyerabend, für alle Begriffstypen. Damit wird nicht nur die Bedeutung von abstrakten theoretischen Begriffen durch deren Einbettung in Aussagenzusammenhänge bestimmt, sondern auch die von elementareren Begriffstypen, wie bspw. Begriffe der Erfahrung (vgl. Feyerabend, 1965, S. 180; Perovich, 1981, S. 79; Sankey, 1994, S. 8 f.). In der Sprache unseres Bezugsrahmens entfernt sich Feyerabend dementsprechend entlang der Dimension Begriffstheorie weiter von der augustinischen Begriffstheorie als dies bei Kuhn der Fall ist.

Nimmt man diese zwei Punkte ernst und wendet gleichzeitig unseren Bezugsrahmen an, dann müsste die Inkommensurabilitätsrelation für die Theorieanlage von Feyerabend eine stärkere Ausprägung aufweisen als bei Kuhn. Der Grund liegt darin, dass sich die Position von Feyerabend auf einer – in Pfeilrichtung – höheren Isolinie befindet, als diejenige von Kuhn.[16] In der Tat stellt bei Feyerabend Inkommensurabilität ein globales und nicht nur lokales Problem dar (vgl. Carrier, 2001, S. 83). Probleme der Verständigung zwischen Vertretern von inkommensurablen Theorien erstrecken sich laut Feyerabend auf ganze Theoriesysteme und nicht wie bei Kuhn ausschliesslich auf kleine Untermengen von Aussagen (vgl. Hoyningen-Huene, 1990, S. 488 f.). Damit konstatiert Feyerabend, dass die Problemklasse, die mit der Inkommensurabilität verbunden ist, grösser ist als diejenige, die Kuhn bereit ist zuzugeben.

Zusammenfassend können wir festhalten, dass die von uns vorgeschlagenen Eskalationsdimensionen für die Inkommensurabilitätsrelation sinnvolle Strukturierungselemente darstellen. Die Betrachtung der Theorieanlagen von Feyerabend und Kuhn hat darüber hinaus deutlich gemacht, dass die beiden identifizierten Bedingungen zusammengenommen plausibel als hinreichendes Kriterienpaket für das Zustandekommen der Inkommensurabilitätsrelation angesehen werden können.

[16] Siehe erneut Abbildung 4

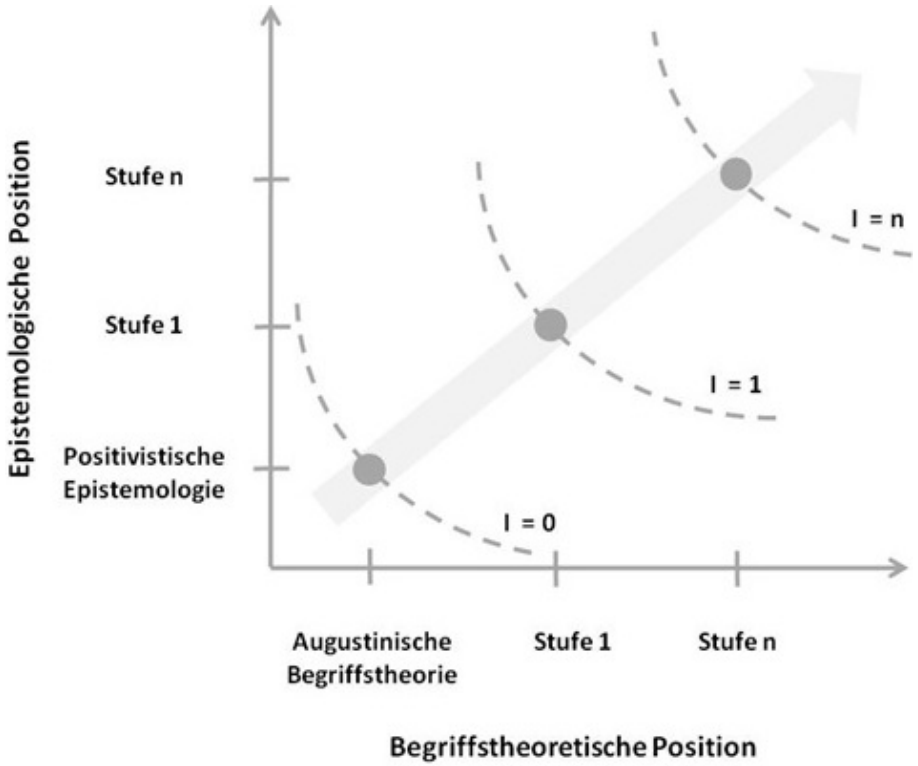

Abbildung 4: Eskalationsdimensionen zur Inkommensurabilitätsrelation (Quelle: eigene Darstellung).

2.2.4. Zwischenfazit – Der entfaltete metatheoretische Bezugsrahmen

Die Identifikation von vier zentralen Elementen des Kuhnschen Begriffs von Inkommensurabilität erlaubte es eingangs Inkommensurabilität als zweigliedrigen, relationalen Begriff zu definieren, mit dem im Kontext von Theoriewahlsituationen bestimmte Typen von Problemen verbunden sind. Dieses Verständnis von Inkommensurabilität wurde anhand einer metatheoretischen Perspektive weiter analysiert. Dabei galt die Aufmerksamkeit nicht dem paarweisen Vergleich von konkreten Aussagensystemen im Hinblick auf Inkommensurabilitätsverhältnisse. Stattdessen wurden die Grundlagen der Inkommensurabilitätsrelation beleuchtet. Die Analyse förderte eine begriffstheoretische und eine epistemologische Bedingung zu Tage, die einzeln notwendig und zusammen hinreichend dafür sind, dass eine derartige Relation auftritt. Vor dem Hintergrund der zwei Bedingungen „nicht-augustinische Begriffstheorie“ und „nicht-positivistische Epistemologie“ lässt sich Inkommensurabilität als graduelles Phänomen begreifen. Die Ausprägung der Probleme der Verständigung und epistemischen Referenz hängt davon ab, welche begriffstheoretischen und epistemologischen Zugeständnisse man in seinem Theoriekonstrukt eingeht. Die nachfolgende Abbildung 5 visualisiert den vollständigen und somit entfalteten metatheoretischen Bezugsrahmen.

Dieser Bezugsrahmen gibt dem Begriff der Inkommensurabilität eine Form, die es erlaubt, die organisationstheoretischen Konsequenzen dieser Thematik sichtbar zu machen. Damit ist die erste Forschungsfrage beantwortet.

Nachfolgend werden die organisationstheoretischen Implikationen exemplarisch anhand der evolutionären Organisationstheorie dargelegt. Der Fokus liegt dabei auf der detaillierten Ausarbeitung der Problemfelder, die sich für diese Organisationstheorie skizzieren lassen. Die bisher nur auf einer abstrakten Ebene gewonnen Einsichten erlangen dadurch eine konkrete und zugängliche Form.

3. Die evolutionäre Organisationstheorie vor dem Problem der Inkommensurabilität

Gegenstand dieses Kapitels bildet die Diskussion der organisationstheoretischen Implikationen, die sich aus dem erarbeiteten Verständnis von Inkommensurabilität ergeben. Unsere metatheoretische Perspektive erlaubt dabei die Erläuterung der Inkommensurabilität in Bezug auf eine Theorie. Indem wir das wissenschaftstheoretische Fundament der evolutionären Organisationstheorie der Münchner Schule[17] genauer betrachten, können wir feststellen, wie ausgeprägt die Inkommensurabilitätsproblematik für diese Theorieanlage ist und welche Konsequenzen für diesen organisationstheoretischen Ansatz gezogen werden müssen. Wie bereits vorangehend angemerkt, besteht das primäre Ziel darin, die in Kapitel 2 gewonnen Einsichten, die rein wissenschaftstheoretisch angelegt sind, zu konkretisieren und für die Organisationstheorie exemplarisch fruchtbar zu machen.

[17] Nachfolgend EOT abgekürzt.

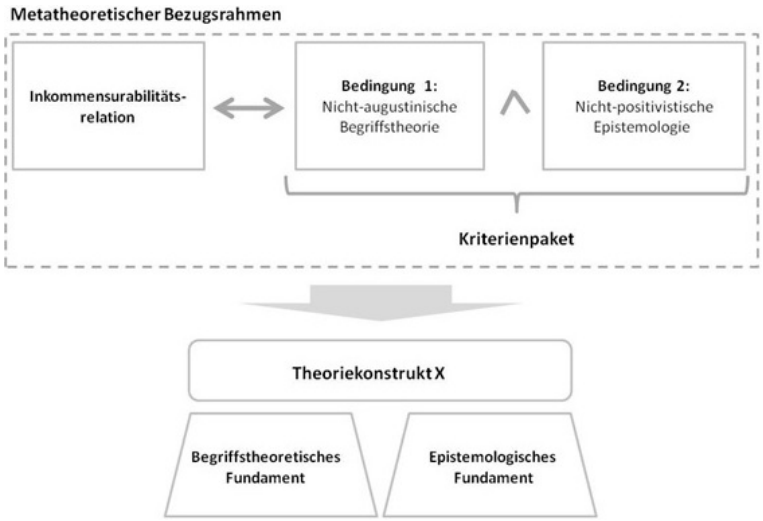

Abbildung 5: Entfalteter metatheoretischer Bezugsrahmen zur Inkommensurabilitätsrelation (Quelle: eigene Darstellung; Die zwei aussagenlogischen Zeichen "↔" und "∧", die in Abbildung 5 vorkommen, werden in Online-Appendix A erläutert.).

Das Kapitel ist wie folgt gegliedert. In *Abschnitt 3.1* wird begründet, warum gerade die EOT als Transferobjekt aus der Vielzahl organisationstheoretischer Ansätze ausgewählt wird. *Abschnitt 3.2* umfasst eine Skizze zentraler Bausteine der EOT. Durch die Identifikation und Erläuterung von fünf zentralen Elementen lässt sich zum einen ein Bild dieser Organisationstheorie gewinnen, zum anderen bilden diese Elemente die Ansatzpunkte, um in *Abschnitt 3.3* die Implikationen der Inkommensurabilitätsproblematik für diese Theorieanlage darzustellen. Die Implikationen werden im Zuge der Erläuterung sowohl auf inter- wie auch intratheoretischer Ebene diskutiert.

3.1. Die evolutionäre Organisationstheorie im Fokus

Die EOT wird seit Mitte der siebziger Jahre am Seminar für strategische Unternehmensführung der LMU München federführend durch Kirsch und seine Schüler entwickelt (vgl. Kirsch, 1992, S. 1). Das bis dato entstandene Projekt generierte einen beträchtlichen Korpus an Literatur. Dieser umfasst eine Reihe von Arbeitstexten (Kirsch, 1998; Kirsch et al., 1999a; Kirsch und Guggemos, 1999b; Kirsch und Weber, 1999c,d; Kirsch und Heeckt, 2001), vier abgeschlossene Buchprojekte (Kirsch, 1992, 1996; Kirsch et al., 2007, 2009) sowie zahlreiche Dissertationen. In der nachfolgenden Analyse werden die Arbeitstexte und Buchprojekte als Textgrundlage verwendet. Bevor diese Organisationstheorie in Abschnitt 3.2 näher beleuchtet wird, ist die Frage zu thematisieren, warum es sinnvoll ist, gerade diesen Ansatz für die Anwendung des metatheoretischen Bezugsrahmens auszuwählen.

Der Beantwortung dieser Frage sind zwei Vorbemerkungen vorauszuschicken. Zum einen kann grundsätzlich jeder organisationstheoretische Ansatz als Transferobjekt dienen. Der in Kapitel 2 erarbeitete Bezugsrahmen, der ein Kriterienpaket bereithält, um die Ausprägung der Inkommensurabilitätsrelation für eine Theorieanalage zu bestimmen, kann immer dann angewandt werden, wenn Klarheit über die begriffstheoretischen und epistemologischen Annahmen der in Frage stehenden Organisationstheorie besteht. Diese Grundannahmen müssen vom organisationstheoretischen Ansatz nicht explizit artikuliert, sondern können auch durch eine Analyse zentraler Bausteine der Theorie, insbesondere der propagierten Forschungsmethodik, erarbeitet werden. Zum anderen hat die Anwendung des metatheoretischen Bezugsrahmens auf die EOT primär exemplarischen Charakter. Die Ausführungen in den Abschnitten 3.2 und 3.3 weisen neben der inhaltlichen eine ebenso wichtige funktionale Ebene auf. Anhand der Diskussion der EOT lässt sich die Methodik aufzeigen, die im Rahmen von Kapitel 2 entwickelt wurde und aus unserer Sicht für die Auseinandersetzung mit der Inkommensurabilitätsproblematik im Feld der Organisationstheorie fruchtbar ist. Es ist an dieser Stelle zu betonen, dass die exemplarische Diskussion der EOT neben diesem funktionalen Aspekt auf der inhaltlichen Ebene keine generalisierbaren Elemente enthält. Die Diskussion der Inkommensurabilitätsproblematik muss für jeden organisationstheoretischen Ansatz inhaltlich einzeln durchgeführt werden.

Wenden wir uns nun der Beantwortung der Ausgangsfrage zu. Die EOT wurde aus vier Gründen als Transferobjekt ausgewählt.

Erstens ist die EOT bezüglich ihres wissenschaftstheoretischen Fundaments eine reflektierte Theorie. Dies äussert sich zum einen darin, dass Teile der sozialwissenschaftlichen und philosophischen Diskussion den Ausgangspunkt des Theorieprojekts bilden (vgl. Kirsch et al., 1999a, S. 8). Zum anderen findet sich der Anspruch mit der EOT einen organisationstheoretischen Rahmen in der Form eines Sprachspiels für organisationale Phänomene bereitzustellen (vgl. Kirsch et al., 2007, S. 26 ff., 79). Dies unterstreicht wiederum das Vorhaben der EOT, Grundlegendes zur organisationstheoretischen Debatte beizutragen. Damit ist die EOT ein geeigneter Kandidat für die Anwendung des metatheoretischen

Bezugsrahmens, der als Einstiegsfläche die begriffstheoretischen und epistemologischen Grundannahmen einer Theorieanlage nutzt.

Zweitens ist eine pluralistische Perspektive das Leitmotiv der EOT (vgl. Kirsch, 1996, S. 323; Kirsch et al., 2007, S. 156 ff.). Indem die EOT als Rahmentheorie angelegt ist, die eine Vielzahl organisationaler Ansätze beherbergt und gleichzeitig der Interaktion mit diesen Ansätzen eine zentrale Rolle bei der Weiterentwicklung der Theorie einräumt (vgl. Kirsch et al., 2007, S. 81 ff.), tritt das Thema der Theorievergleiche und damit die Problematik der Inkommensurabilität auf die Agenda. Es ist dementsprechend angebracht, für ein solches Theorieprojekt sauber zu analysieren, welche Rolle Inkommensurabilität darin spielt.

Drittens ist die EOT ein fluides Theoriegebäude in doppeltem Sinne. Die EOT wird zum einen nicht als statische Theorie, sondern als veränderbarer Theorieprozess konzipiert (vgl. Kirsch und Weber, 1999c, S. 85). Wie in Abschnitt 3.2 weiter auszuführen ist, begreift sich die EOT als evolvierende Theorieanlage, die sich mit ihrem Objektbereich – Organisationen – über die Zeit weiterentwickelt (vgl. Kirsch, 1998, S. 123 f.). Zum anderen erlaubt es diese Flexibilität in der Theorieanlage kontinuierlich neue theoretische Einsichten zu integrieren und Anpassungen vorzunehmen (vgl. Kirsch und Weber, 1999c, S. 106 ff.). Durch diese fluide Theoriekonzeption wird die EOT zum idealen Transferobjekt, da die erarbeiteten Implikationen der Inkommensurabilität als Konstruktionshinweise für die EOT aufgefasst werden können.

Viertens wird das Thema Inkommensurabilität innerhalb der EOT an zahlreichen Orten thematisiert (vgl. exemplarisch Kirsch, 1998, S. 10 ff., 30, 36 f.; Kirsch und Weber, 1999d, S. 51, 157; Kirsch und Heeckt, 2001, S. 14). Was bei dieser Thematisierung jedoch fehlt, ist eine Reflexion der Inkommensurabilität vor dem Hintergrund ihrer begriffstheoretischen und epistemologischen Grundlagen. An dieser Stelle kann der metatheoretische Bezugsrahmen einen sinnvollen Beitrag leisten.

Widmen wir uns in einem nächsten Schritt der Charakterisierung der EOT.

3.2. Skizze der evolutionären Organisationstheorie

Die Bemerkungen zur Textgrundlage der EOT haben bereits angezeigt, dass es sich bei dieser Organisationstheorie um ein über die Zeit gewachsenes, inhaltlich umfangreiches und ausdifferenziertes Projekt handelt. Eine Rekonstruktion der EOT kann dementsprechend im Rahmen dieser Untersuchung nur die Form einer Skizze annehmen. Die Skizze umfasst jene Aspekte, welche von den Autoren als zentrale Bausteine der EOT bezeichnet werden und zugleich fruchtbare Reflexionsfelder darstellen, um in Abschnitt 3.3 die Implikationen der Inkommensurabilität sichtbar zu machen.

Abbildung 6 gibt einen Überblick über die fünf Elemente der EOT, die im Rahmen der Skizze in kompakter Form diskutiert werden.

Beginnen wir die Erläuterungen mit dem wissenschaftstheoretischen Fundament der EOT. Darunter werden hier die begriffstheoretischen und epistemologischen Annahmen verstanden, die der Theorieanlage zu Grunde liegen. Die EOT geht erstens von einer Gebrauchstheorie der Bedeutung aus, die dem späten Wittgenstein entlehnt ist (vgl. exemplarisch Kirsch, 1998, S. 185). Sprachliche Bedeutung, so die Autoren, wird durch Handlungen, bei denen sprachliche Elemente in Handlungszusammenhänge verwoben werden, erzeugt (vgl. Kirsch et al., 1999a, S. 296, 300). Diese Handlungs- und Sprachzusammenhänge werden als Sprachspiele bezeichnet (vgl. Kirsch und Heeckt, 2001, S. 12). Die Summe derartiger Sprachspiele, die in unserer persönlichen Erfahrungswelt vorkommen, ist Teil einer spezifischen Lebensform (vgl. Kirsch, 1992, S. 9; Kirsch und Guggemos, 1999b, S. 16).[18] Ein Beispiel kann diese Punkte illustrieren. Der Begriff „Tisch“ hat sprachliche Bedeutung, weil er in bestimmten Sprachspielen vorkommt und darin jeweils Anknüpfungspunkte für Handlungen ermöglicht. So verstehen wir bspw. die Aussage „Setz dich an den Tisch!“ als Aufforderung und handeln entsprechend.

Zweitens findet sich in der EOT die These des radikalen Konstruktivismus (vgl. exemplarisch Kirsch et al., 1999a, S. 362). Aus erkenntnistheoretischer Perspektive betrachtet, besagt diese These, dass Wissen als aktive, selbstbezügliche und damit radikal subjektive Konstruktion eines Beobachters aufzufassen ist (vgl. ebd., S. 286 ff.; Kirsch und Weber, 1999c, S. 113). Damit bricht diese Position mit dem naiven Realismus, der menschliche Erkenntnisleistung als passives Abbilden einer beobachterunabhängigen Wirklichkeit auffasst (vgl. Kirsch und Weber, 1999c, S. 95). Die These des radikalen Konstruktivismus ist jedoch keine idealistische Position. Wirklichkeitskonstruktionen sind zwar beobachterabhängig, aber dennoch nicht ohne Bezug zu einer externen Umwelt. Die externe, nur indirekt erkennbare Wirklichkeit bildet den Prüfstein für unsere Wirklichkeitskonstruktionen, die sich als viabel – als für unsere Zwecke nützlich – erweisen müssen (vgl. Kirsch, 1998, S. 122; Kirsch et al., 1999a, S. 61 ff.).

Soweit die Ausführungen zum wissenschaftstheoretischen Fundament. Wenden wir uns dem *Theorieverständnis* zu, dass der EOT zu Grunde liegt. Die EOT ist keine statische Theorie, sondern ist als Theorieprozess angelegt (vgl. Kirsch et al., 2009, S. 37 f.). Darunter ist eine zeitliche Abfolge verschiedener Formen der EOT zu verstehen (vgl. Kirsch, 1998, S. 238). Es lassen sich drei Haupttreiber identifizieren, die als Motor für die stetige Veränderung der EOT dienen. Erstens geht die EOT von einer Evolution ihres Objektbereichs aus (vgl. Kirsch et al., 2009, S. 281). Organisationen sind soziale Phänomene, die sich aufgrund der Kreativität und Innovationskraft ihrer Akteure ständig verändern (vgl. Kirsch et al., 2007, S. 36). Eine Organisationstheorie, die diesem Wandel gerecht werden will, ist nicht als statisches Aussagensystem, sondern als fluides Theoriegebäude zu konzipieren (vgl. Kirsch und Weber, 1999c, S. 106 f.). Zweitens ist die EOT als Rahmentheorie angelegt (vgl. Kirsch et al., 2007, S. 83). Neben Aussagen über Organisationen finden

[18] Die Thematik der Lebensformen wird unter dem Stichwort „Akteurs- und Problemverständnis“ weiterführend erläutert.

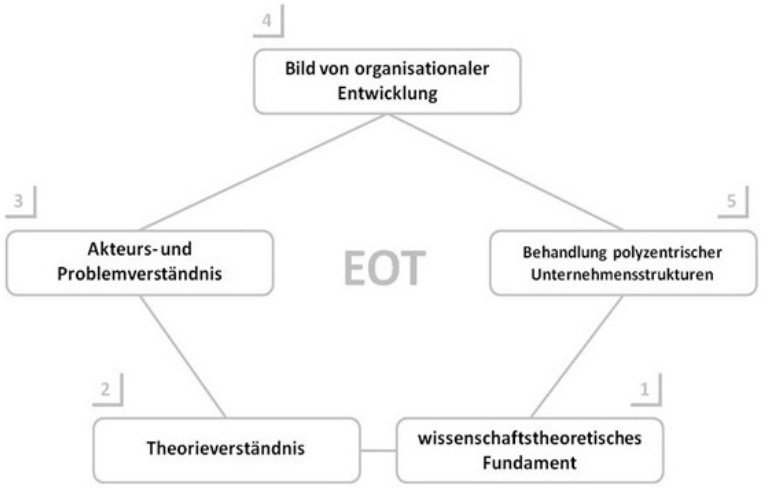

Abbildung 6: Fünf zentrale Elemente der EOT (Quelle: eigene Darstellung).

sich auch metatheoretische Aspekte in der EOT, welche deren Beziehung zu anderen Organisationstheorien behandeln (vgl. ebd.). Andere organisationstheoretische Ansätze werden zugleich als Partner und Konkurrenten verstanden. Ihre Einsichten werden in die EOT übernommen, resp. kritisch reflektiert und in modifizierter Form integriert (vgl. ebd., S. 81). Dieses Beziehungsgeflecht zwischen der EOT und konkurrierenden organisationstheoretischen Ansätzen zwingt die EOT zur ständigen Anpassung. Drittens ist die EOT selbstreflexiv. Sie geht davon aus, dass sie selbst in ihrem Objektbereich vorkommt (vgl. Kirsch und Weber, 1999d, S. 184). Indem die EOT Wissenschaftler als relevante Akteure im Umfeld einer Organisation identifiziert, gehören die Ergebnisse betriebswissenschaftlicher Forschung per definitionem in den Gegenstandsbereich der EOT (vgl. Kirsch, 1996, S. 56). Darüber hinaus beziehen die Akteure der Organisation die EOT in ihre Reflexionen mit ein und verändern dadurch ihre Wahrnehmungen der organisationalen Phänomene (vgl. Kirsch, 1998, S. 47).[19] Die so entstehende wechselseitige Beziehung zwischen Theorieentwicklung und Anwendung ist wiederum Ursache für die Dynamik der EOT.

Betrachten wir in einem nächsten Schritt das *Akteurs- und Problemverständnis* der EOT. Die Gruppe der organisationstheoretisch relevanten Akteure, die sich in der EOT findet, ist in zwei Hinsichten spezifisch ausgeprägt. Zum einen gehören neben den „üblichen" Akteuren einer Organisation (Mitarbeiter und direkt gekoppelte Stakeholder wie Lieferanten und Kapitalgeber) auch Analysten, Berater und wissenschaftliche Gemeinschaften zum Gegenstandsbereich der EOT (vgl. Kirsch et al., 2009, S. 281). Zum anderen werden diese internen und externen Akteure als Beobachter konzipiert (vgl. Kirsch et al., 1999a, S. 117). Diese Beobachter konstruieren im Prozess der Beobachtung organisationaler Phänomene (bspw. die Übernahme und Integration einer Konkurrenzfirma) jeweils ihre spezifische organisationale Wirklichkeit. Der Inhalt dieser Wirklichkeitskonstruktion wird durch die Lebenswelt eines Beobachters geprägt (vgl. ebd., S. 133; Kirsch und Heeckt, 2001, S. 12). Lebenswelt ist ein Sammelbegriff für Sprachspiele, Wissensformen und Handlungsweisen, in denen sich jeweils ein Beobachtertyp (bspw. der des Beraters) oder eine Untergruppe davon bewegt (vgl. Kirsch und Guggemos, 1999b, S. 15; Kirsch und Heeckt, 2001, S. 12). Die EOT betont, dass es zwischen diesen Beobachtertypen auch zu wechselseitigen Beobachtungen kommt (vgl. Kirsch et al., 2009, S. 281).

Dieses Bild von organisationalen Akteuren prägt in hohem Mass das Problemverständnis der EOT. Probleme, die in Organisationen auftreten, sind gemäss den Autoren in aller Regel „Multi-Kontext-Probleme" (Kirsch, 1998, S. 10), da mit einem Problem stets unterschiedliche Akteure verbunden sind und somit unterschiedliche Wirklichkeitskonstruktionen miteinander in Berührung stehen (vgl. ebd., S. 12; Kirsch et al., 2007, S. 85). Konsequenterweise propagiert die EOT in Bezug auf organisationale Probleme eine Strategie der „echten Komplexitätsbejahung" (Kirsch, 1998, S. 14). Gemäss dieser Herangehensweise ist die Pluralität der Wirklichkeitskonstruktionen der betroffenen Akteure ernst zu nehmen. Dies geschieht, indem eine reflexive Haltung eingenommen wird, die bestrebt ist, die beteiligten Lebenswelten aus der Binnenperspektive zu verstehen und gemeinsam neue Kontexte zu erzeugen, welche die Basis für Verständigung bilden (vgl. Kirsch et al., 1999a, S. 31 ff., 363; Kirsch et al., 2007, S. 87).

Soweit die Ausführungen zum Akteurs- und Problemverständnis der EOT. Nachfolgend wird das *Bild von organisationaler Entwicklung* skizziert, welches sich in der EOT findet. Wie bereits erläutert, geht die EOT davon aus, dass Organisationen nicht statische, sondern dynamische Objekte darstellen. Die Organisationen sind in diesem Sinne mit einer offenen Zukunft konfrontiert (vgl. Kirsch et al., 2007, S. 36, 88; Kirsch, 1998, S. 118). Gemäss der EOT sind Organisationen dieser Offenheit ihres weiteren Entwicklungspfades nicht passiv ausgeliefert, sondern in der Lage, auf die Entwicklun-

[19] Dieser Aspekt wird unter dem Stichwort „Akteurs- und Problemverständnis" weiter ausgeführt (siehe direkt nachfolgend).

gen zu reagieren und diese bedingt zu steuern (vgl. Kirsch, 1992, S. 331; Kirsch et al., 2009, S. 286). Es stellt sich nun die Frage, ob es innerhalb einer Organisation eine Struktur gibt, welche den Umgang mit zukünftigen, unvorhersehbaren Ereignissen anleitet. Der EOT zufolge ist ein Sinnmodell, das während einer bestimmten Zeitspanne innerhalb einer Organisation wirksam ist, eine derartige Struktur (vgl. Kirsch und Heeckt, 2001, S. 244). Sinnmodell ist ein Sammelbegriff für all jene Inhalte, die festlegen, was Sinn und Zweck einer Organisation ist und damit handlungsanleitenden Charakter in Bezug auf Problemdefinitionen, Situationsbeschreibungen und Lösungsentwicklungen innerhalb des organisationalen Alltags haben (vgl. Kirsch et al., 2009, S. 287). Indem diese Sinnmodelle im organisationalen Alltag verankert werden, sind sie für die gesamte Organisation prägend (vgl. Kirsch und Heeckt, 2001, S. 244 f.). Die EOT unterscheidet paradigmatisch zwischen drei Sinnmodellen, in der die Menge organisational relevanter Stakeholder jeweils unterschiedlich umfassend ist. Mit dem Instrumentalmodell ist die Sichtweise verbunden, dass Organisationen primär als Mittel zur Erfüllung der Interessen ihrer primären Akteuren, bspw. der Eigen- und Fremdkapitalgebern zu betrachten sind (vgl. Kirsch et al., 2007, S. 91). Im Überlebens- resp. Bestandsmodell dominiert die Vorstellung, dass Organisationen in ein Geflecht wechselseitiger Austauschbeziehungen eingebunden sind und dieses aufrechterhalten müssen, um ihr Fortbestehen sichern zu können (vgl. ebd., S. 91 f.). Schliesslich betont das Fortschrittsmodell, dass im Zentrum organisationalen Handelns die Befriedigung der Interessen derjenigen Akteure steht, die direkt oder nur indirekt von organisationalen Aktivitäten betroffen sind (vgl. ebd., S. 92). Es ist zu betonen, dass Sinnmodelle jeweils partiell übernommen werden können. So können bspw. in einer Organisation sowohl das Sinnmodell des zweiten Typs wie auch bereits Elemente des Dritten aktiv sein (vgl. Kirsch, 1992, S. 325).

Das Bild von organisationaler Entwicklung, welches die EOT zeichnet, ist direkt über diese Sinnmodelle zu erläutern. Organisationale Entwicklung bedeutet eine Höherentwicklung (vgl. Kirsch, 1998, S. 19). Die Höherentwicklung erfolgt, indem Organisationen die unterschiedlichen Sinnmodelle in der obigen Reihenfolge durchlaufen und damit die organisational relevanten Vorgaben im Hinblick auf Situationsbeschreibungen, Problemdefinitionen und Lösungsentwicklungen anpassen (vgl. Kirsch, 1998, S. 19 f.; Kirsch und Heeckt, 2001, S. 246).

Wenden wir uns abschliessend der *Behandlung polyzentrischer Unternehmensstrukturen* durch die EOT zu. Polyzentrische Unternehmensstrukturen entstehen dann, wenn eine Organisation entlang mindestens einer der zwei folgenden Dimensionen Veränderungen erfährt (vgl. Kirsch et al., 2009, S. 291; Kirsch, 1998, S. 97 f.). Zum einen kann eine Organisation Verbindungen mit übrigen Unternehmen eingehen. Darunter fallen bspw. Fusionen, Akquisitionen oder Mitgliedschaften in Wirtschaftsverbänden (vgl. Kirsch, 1996, S. 249). Zum anderen kann eine Organisation ihr Tätigkeitsfeld internationalisieren, so bspw. wenn ein in der Schweiz tätiges Unternehmen in Südostasien neue Marktpotentiale erschliesst (vgl. Kirsch, 1998, S. 97; Kirsch et al., 1999a, S. 119). Die EOT hebt innerhalb dieser Unternehmensstrukturen insbesondere drei Aspekte hervor. Erstens sind polyzentrische Handlungsstrukturen nicht durch das klassische Modell der Hierarchie abbild- und begreifbar (vgl. Kirsch, 1996, S. 249). Stattdessen ist dieses Beziehungsnetz durch eine Vielzahl mehr oder weniger eigensinniger Partialzentren geprägt (vgl. ebd.). Darunter fallen dezentrale institutionelle Stellen (wie bspw. Geschäftsleitungen von Tochterunternehmen), die eine Führungsfunktion übernehmen (vgl. Kirsch et al., 1999a, S. 196). Je stärker sich diese Stellen zentralistischen Führungsimpulsen (bspw. einem konzernweiten Reportingstandard) widersetzen, als desto eigensinniger gelten sie innerhalb der EOT (vgl. ebd., S. 197). Zweitens werden die Partialzentren als organisationale Akteure und damit als Beobachter verstanden (vgl. Kirsch, 1996, S. 251). Zwischen den Partialzentren finden wechselseitige Beobachtungen statt, die wiederum die Handlungen der Partialzentren beeinflussen können (vgl. ebd.). Drittens lassen sich innerhalb polyzentrischer Unternehmensstrukturen jeweils im Umfeld von Partialzentren Partialsysteme identifizieren. Diese Systeme sind dadurch gekennzeichnet, dass die darin vorkommenden organisationalen Handlungen weitgehend durch die Führungsimpulse des zugehörigen Partialzentrums bestimmt werden (vgl. ebd., S. 255). Die EOT nimmt an, dass die Partialsysteme jeweils eigene Lebenswelten ausbilden, in denen unterschiedliche Sprachspiele und Handlungsmuster dominieren (vgl. ebd.). Je nachdem, wie stark sich diese Lebenswelten zwischen den Partialsystemen unterscheiden, spricht die EOT von einem Polyzentrismus im starken oder schwachen Sinne (vgl. Kirsch, 1998, S. 100 f.; Kirsch et al., 1999a, S. 206). Eine breit diversifizierte Holding mit heterogen Geschäftsfeldern ist ein Beispiel für die erste, ein Kartell in einem wettbewerbsarmen Markt ein Beispiel für die zweite Ausprägung (vgl. Kirsch, 1996, S. 274).

Damit sind die Ausführungen zu den fünf, für unsere Untersuchung relevanten, Elemente der EOT abgeschlossen. Die EOT ist in den Grobzügen skizziert und für den Zweck unserer weiterführenden Analyse in hinreichender Tiefe präsent.

3.3. Implikationen für die evolutionäre Organisationstheorie

Gegenstand dieses Abschnitts ist die Ausarbeitung der Implikationen, die sich aus dem Inkommensurabilitätsbegriff ergeben, der in Kapitel 2 hergeleitet wurde. Die Implikationen werden auf drei Ebenen dargelegt. Diese drei Ebenen lassen sich je durch eine pointierte Frage erfassen. In einem ersten Schritt (*Abschnitt 3.3.1*) wird die Ausprägung der Inkommensurabilitätsrelation innerhalb der EOT untersucht. Die zentrale Frage, welche bei der Diskussion dieser Ebene im Vordergrund steht, lautet: Welche Ausprägung der Inkommensurabilität muss die EOT akzeptieren, wenn sie ihr wissenschaftstheoretisches Fundament ernst nimmt? Davon ausgehend wird in einem nächsten Schritt (*Abschnitt 3.3.2*) erläutert, wie die Implikationen der Inkommensurabilität für die EOT auf intertheoretischer Ebene aussehen. Anleitend für diese Ebene der Betrachtung ist die Frage, welche Folgen Inkommensurabilität für die Beziehung der EOT mit konkurrie-

renden organisationstheoretischen Ansätzen hat. Abschliessend (*Abschnitt 3.3.3*) wenden wir uns – wiederum basierend auf der Beantwortung der ersten Frage – den intratheoretischen Implikationen für die EOT zu. Die anleitende Frage in Bezug auf diese Ebene ist Folgende: Was passiert mit den Aussagen der EOT über organisationale Phänomene, wenn wir die Inkommensurabilitätsthematik ernst nehmen?

Abbildung 7 veranschaulicht die drei Ebenen der Diskussion. Sie zeigt zudem an, welche Elemente der EOT für welche Ebene der Implikationen relevant sind.

3.3.1. Ausprägung der Inkommensurabilitätsrelation

Um die Frage nach der Ausprägung der Inkommensurabilitätsrelation innerhalb der EOT zu klären, können wir auf den metatheoretischen Bezugsrahmen zurückgreifen, den wir in Kapitel 2 entwickelt haben.

Das Kriterienpaket, welches den Kern unseres Bezugsrahmens[20] ausmacht, lässt sich direkt auf das wissenschaftstheoretische Fundament der EOT anwenden. Die zentrale Aussage des Kriterienpakets sei an dieser Stelle noch einmal wiederholt: Findet sich eine Theorieanlage, die eine nicht-augustinische Begriffstheorie und eine nicht-positivistische Epistemologie voraussetzt, dann muss diese Theorieanlage Inkommensurabilität und die damit verbundenen Probleme ernst nehmen. Die weitere Argumentation ist damit vorgezeichnet. In einem ersten Schritt gilt es zu prüfen, ob die Gebrauchstheorie der Bedeutung eine nicht-augustinische Begriffstheorie darstellt. Analog ist für die These des radikalen Konstruktivismus zu eruieren, ob es sich dabei um eine nicht-positivistische Epistemologie handelt.

Fokussieren wir als erstes auf die *Gebrauchstheorie der Bedeutung*. Wie in den Ausführungen zum wissenschaftstheoretischen Fundament der EOT dargelegt, geht diese Bedeutungstheorie davon aus, dass Bedeutung entsteht, indem sprachliche Ausdrücke in Handlungszusammenhänge verwoben werden. Diese sogenannten Sprachspiele bilden den Bezugspunkt, wenn man über Sprache und Verständigungsprobleme nachdenkt. Für unsere Untersuchung ergeben sich zwei relevante Fragen: Findet sich innerhalb dieses Verständnisses von sprachlicher Bedeutung das zentrale Merkmal der augustinischen Begriffstheorie? Falls Nein; Wie weit ist die Gebrauchstheorie von der augustinischen Begriffstheorie entfernt?

Die Gebrauchstheorie der Bedeutung wurde von Wittgenstein explizit in Abgrenzung zu einem augustinischen Bild von Sprache konzipiert (vgl. Wittgenstein, 1999, S. 238 f.; Glock, 1996, S. 41). Worin diese Abgrenzung besteht, wird erkennbar, wenn wir uns erneut das zentrale Merkmal der augustinischen Begriffstheorie vergegenwärtigen. Dieses besteht in dem Postulat einer intersubjektiv zugänglichen Ebene von bedeutungsstiftenden Objekten. Es ist diese Ebene, welche sprachliche Bedeutung erklärt und zugleich deren Kontextinvarianz sicherstellt. Dieser Versuch, Bedeutung über eine Ebene intersubjektiv zugänglicher Objekte zu erklären, wird von Wittgenstein zurückgewiesen (vgl. exemplarisch Wittgenstein, 1999, S. 239, 252). Indem er zulässt, dass Begriffe in Abhängigkeit ihrer Verwendung in Handlungszusammenhängen unterschiedliche Funktionen übernehmen können (vgl. Wittgenstein, 1999, S. 262; Schulte, 2001, S. 138), bricht er mit dem objektzentrierten Bild sprachlicher Bedeutung. Wir können dementsprechend festhalten, dass die Gebrauchstheorie das zentrale Merkmal der augustinischen Begriffstheorie zurückweist und es sich somit um eine nicht-augustinische Begriffstheorie handelt.

Die Ausführungen zum graduellen Charakter der Inkommensurabilität haben deutlich gemacht, dass man sich mehr oder weniger stark von einem zentralen Punkt einer Position entfernen kann. Es stellt sich nun folgende Frage: Wie weit voneinander entfernt liegen die augustinische und die wittgensteinsche Auffassung von sprachlicher Bedeutung? Diese Frage kann nur relational und nicht durch einen absoluten Wert beantwortet werden. Aus diesem Grund bietet es sich an, die Gebrauchstheorie zu den Bedeutungstheorien von Kuhn und Feyerabend in Beziehung zu setzen. Wir haben gesehen, dass sich sowohl bei Kuhn, wie auch bei Feyerabend eine Kontexttheorie der Bedeutung finden lässt. Feyerabend öffnet seine Variante der Kontexttheorie für eine grössere Klasse von Begriffen als dies Kuhn tut. Vergleicht man die Gebrauchstheorie mit der Kontexttheorie in der Version Feyerabends, so ist erkennbar, dass die Gebrauchstheorie entlang der begrifflichen Dimension weiter von der augustinischen Ausgangsposition entfernt ist, als dies bei Feyerabends Kontexttheorie der Fall ist. Innerhalb der Gebrauchstheorie haben nicht nur die Vielfalt an sprachlichen Aussagenzusammenhängen eine bedeutungsstiftende Funktion, sondern auch alle Schattierungen menschlichen Handelns (vgl. Glock, 1996, S. 376 f.). Damit steigt die Möglichkeit von Bedeutungsunterschieden über unterschiedliche Handlungskontexte hinweg an. Die Vorstellung von kontextunabhängiger und intersubjektiv geteilter Bedeutung wird damit eliminiert. Es lässt sich demzufolge konstatieren, dass es sich beim begriffstheoretischen Fundament der EOT im Vergleich zu demjenigen von Kuhn und Feyerabend um ein Pluralistischeres handelt.

Soweit die Ausführungen zur Gebrauchstheorie der Bedeutung. Verfahren wir zweitens mit der *These des radikalen Konstruktivismus* analog. Wie bereits in den Ausführungen zum wissenschaftstheoretischen Fundament der EOT dargelegt, geht diese Position davon aus, dass Wissen eine aktive, beobachterabhängige und viable Konstruktion darstellt. Wiederum stehen aus Sicht unserer Untersuchung zwei relevante Fragen an: Findet sich im Kern des radikalen Konstruktivismus das zentrale Merkmal einer positivistischen Epistemologie? Falls Nein; Wie weit ist die These des radikalen Konstruktivismus von der positivistischen Epistemologie entfernt?

Das zentrale Merkmal der positivistischen Epistemologie besteht in einer intersubjektiv zugänglichen Ebene von primitiven Wahrnehmungsinhalten. Diese Wahrnehmungseindrücke bilden nicht nur die Grundlage unserer komplexeren Erfahrungsinhalte, sondern auch den Bezugspunkt, wenn wir eine Rechtfertigung für unsere Wissensansprüche liefern.

[20] Für die graphische Visualisierung siehe erneut Abbildung 5

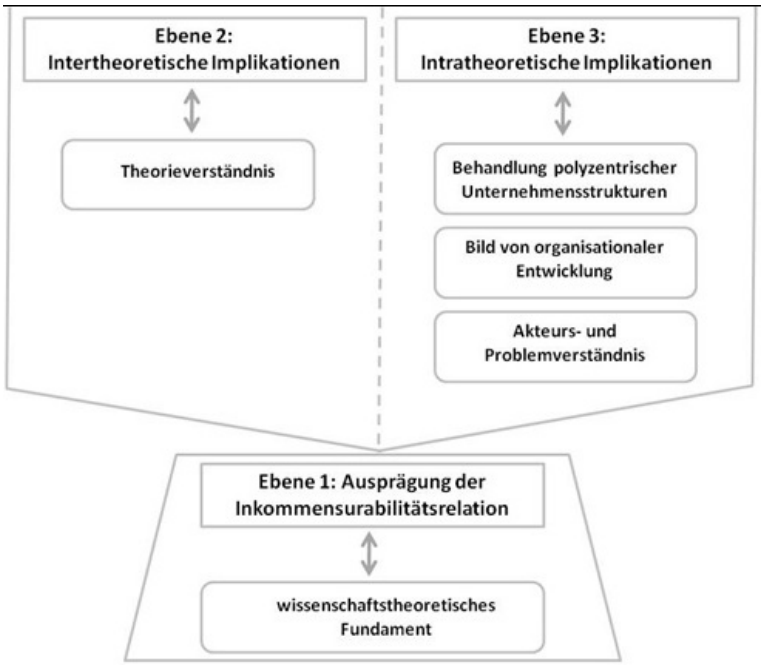

Abbildung 7: Überblick über die Diskussion der Implikationen der Inkommensurabilität für die EOT (Quelle: eigene Darstellung).

Präzise mit dieser Vorstellung bricht der radikale Konstruktivismus. Indem Wissen ein abbildender und somit referentieller Charakter abgesprochen und stattdessen die operationale Geschlossenheit unseres kognitiven Apparats postuliert wird, weist diese Position die These von intersubjektiv teilbaren Wirklichkeitskonstruktionen zurück (vgl. Kirsch, 1992, S. 199 ff., 487 f.; Kirsch, 1996, S. 99 f.; Kirsch, 1998, S. 170, 182; Kirsch et al., 1999a, S. 16, 61 ff.). Wir können demzufolge konstatieren, dass die These des radikalen Konstruktivismus das zentrale Merkmal der positivistischen Epistemologie ablehnt und es sich somit um eine nicht-positivistische Epistemologie handelt.

Vergleichen wir die These des radikalen Konstruktivismus mit den epistemologischen Grundannahmen, die bei Kuhn und Feyerabend zu finden sind. Wir stellten fest, dass Kuhn und Feyerabend die Annahme teilen, dass unsere Wahrnehmung massgeblich durch die begrifflichen Strukturen, die uns zur Verfügung stehen, geprägt wird. Subjektseitigen Faktoren kommt somit eine konstitutive Rolle in der Wahrnehmung zu. Dabei verhaften sie jedoch im neo-kantianischen Bild, welches besagt, dass diese Begriffskategorien intersubjektiv geteilt werden können (so bspw. zwischen den Mitgliedern einer wissenschaftlichen Gemeinschaft). Indem der Konstruktivismus die radikale Subjektivität der Wirklichkeitskonstruktionen zulässt, schwindet die Vorstellung von intersubjektiv geteilten Wahrnehmungsinhalten. Es lässt sich wiederum festhalten, dass das epistemologische Fundament der EOT im Vergleich zu demjenigen von Kuhn und Feyerabend pluralistischer ausgestaltet ist.

An diesem Punkt der Argumentation lässt sich ein *Einwand* vorbringen. Zugegeben, die EOT setzt eine nicht-augustinische und eine nicht-positivistische Epistemologie voraus. Damit entfallen bedeutungsstiftende Objekte und primäre Sinneseindrücke als mögliche Ebenen, die Klarheit im Hinblick auf sprachliche Bedeutung und Wahrnehmungsinhalte sicherstellen. Jedoch ist selbst unter der Voraussetzung der Gebrauchstheorie der Bedeutung und der These des radikalen Konstruktivismus eine intersubjektiv zugängliche Ebene, welche diese Funktionen übernimmt, immer noch denkbar. Eine solche argumentative Strategie findet sich innerhalb der EOT (vgl. exemplarisch Kirsch et al., 1999a, S. 97 ff.). Die Autoren geben zwar zu, dass „vor dem Hintergrund konstruktivistischer Überlegungen (...) insbesondere die Intersubjektivitätsannahme (...) zu relativieren ist“ (ebd., S. 97). Sie betonen aber zugleich, dass unter der Voraussetzung einer konstruktivistischen Epistemologie verständigungsorientiertes Handeln in Form eines relativierten Verständnisses von Intersubjektivität möglich ist (vgl. ebd.). Die Argumentation lautet wie folgt: Setzt man voraus, dass unterschiedliche Beobachter in einem gemeinsamen Wirkungsbereich operieren, dann kann aufgrund einer strukturellen Kopplung der Akteure eine Parallelisierung der Wirklichkeitskonstruktionen entstehen (vgl. ebd., S. 98). Bei dieser Strategie gibt man zwar zu, dass von intersubjektiv geteiltem Wissen oder einem gemeinsamen Vorrat an sprachlicher Bedeutung nicht die Rede sein kann. Gleichzeitig konstatiert man jedoch, dass durch eine Abschwächung der radikalen Subjektivität der Wirklichkeitskonstruktionen eine Annäherung zwischen Beobachtern der gleichen Lebenswelt wahrscheinlich ist (vgl. ebd.). Von einer derartigen Annäherung ist insbesondere bei menschlichen Beobachtern auszugehen, da mit unserer Sprache ein geradezu ideales Werkzeug zur Verfügung steht,

um verständigungsorientierte Parallelisierungen vorzunehmen (vgl. ebd.). Damit ist die Bedrohungslage, die durch die Inkommensurabilität hervorgerufen wird, für die EOT weit weniger akut, als es die vorangegangen Ausführungen nahelegen.

Diese Argumentation hinkt an einer zentralen Stelle. Wie die Autoren reflektiert bemerken, baut diese Strategie auf einer „abgemilderten Version eines radikalen Konstruktivismus“ (ebd.) auf. Nimmt man dagegen die Zugeständnisse der EOT zu einem radikalen Konstruktivismus ernst (vgl. exemplarisch Kirsch et al., 1999a, S. 286; Kirsch und Weber, 1999c, S. 113), dann ist diese Abschwächung kein zulässiger argumentativer Schritt. Der radikale Konstruktivismus hat erhebliche Schwierigkeiten mit der Frage, wie eine Form der Intersubjektivität und damit ein referentieller Bezugspunkt entstehen kann, wenn man die Annahme eines operational geschlossenen kognitiven Apparats des Beobachters ernst nimmt (vgl. Winter, 1999, S. 157). Der Versuch von Kirsch et al. (1999a, S. 98 ff.), der analog zu demjenigen von von Glaserfeld (1985, S. 19 ff.) funktioniert, liefert hier keine befriedigende Lösung. Indem rein spekulativ angenommen wird, dass eine beobachterabhängige Konstruktion von Intersubjektivität möglich ist (vgl. Kirsch et al., 1999a, S. 99), geschieht eine Abschwächung dessen, was den Kern des radikalen Konstruktivismus ausmacht: Die Vorstellung, dass Wirklichkeit eine subjektive Konstruktion darstellt (vgl. hierzu auch Winter, 1999, S. 157 f.).[21]

Basierend auf dieser Feststellung können wir rekapitulierend festhalten: Wann immer in der EOT explizit von Inkommensurabilität die Rede ist oder Inkommensurabilität aufgrund der behandelten Verhältnisse präsent ist, liegt Inkommensurabilität in einem starken Sinne vor. Damit ist die EOT akut mit Problemen der Verständigung und epistemischen Rechtfertigung konfrontiert. Diese zentrale Einsicht veranschaulicht Abbildung 8. Die Abbildung 8 dient uns als graphischer Bezugspunkt für die nachfolgende Diskussion der inter- und intratheoretischen Implikationen der Inkommensurabilität.

3.3.2. Intertheoretische Implikationen der Inkommensurabilität

Mit der Bestimmung der Ausprägung der Inkommensurabilitätsrelation für die EOT wurde die Voraussetzung geschaffen, um die zweite und dritte Ebene der Implikationen[22] zu diskutieren. Wir können in einem ersten Schritt eruieren, worin die Folgen eines starken Begriffs von Inkommensurabilität für das Theorieverständnis der EOT bestehen. Insbesondere steht die Frage im Raum, wie es unter Berücksichtigung der Inkommensurabilität möglich ist, eine Theorie als dynamischen Theorieprozess auszugestalten, in dem explizit konkurrierende organisationstheoretische Ansätze berücksichtigt werden.

Die Präsentation dieser Implikationen ist wie folgt strukturiert. Sowohl die inter- (*Abschnitt 3.3.2*) wie auch die intratheoretischen Implikationen (*Abschnitt 3.3.3*) werden in Form von insgesamt acht pointierten, kritischen Befunden formuliert. Diese sind nachfolgend einzeln aufgeführt und begründet. Dadurch konkretisiert sich die Problemtypologie, die mit der Inkommensurabilitätsrelation verbunden ist.

Die Ausführungen zum Theorieverständnis haben deutlich gemacht, dass die EOT keine statische Theorie im Sinne eines stabilen Aussagensystems ist. Vielmehr handelt es sich bei der EOT um einen Theorieprozess, der seine Eigendynamik drei Haupttreibern verdankt. Zentralster Treiber ist das Beziehungsgeflecht der EOT zu konkurrierenden organisatorischen Ansätzen, durch das alternative theoretische Einsichten zu organisationalen Phänomenen Eingang in die EOT finden. Ausgehend von diesem Theorieverständnis lassen sich zwei kritische Befunde identifizieren.[23]

Die erste intertheoretische Implikation der Inkommensurabilität lautet wie folgt:

> *Befund 1:* Die Konzeption der EOT als Rahmentheorie, die Einsichten von konkurrierenden organisationstheoretischen Ansätzen fruchtbar integrieren kann, ist problematisch.

Sobald die EOT in einen Austausch mit konkurrierenden Ansätzen tritt, entsteht eine relationale Beziehung zwischen zwei Aussagensystemen. Die EOT nimmt Bezug auf Forschungsergebnisse zu oder Konzeptionalisierungen von organisationalen Phänomenen, die in einem konkurrierenden Ansatz erarbeitet wurden. Betrachten wir diesen Inputstrom für die EOT vor dem Hintergrund eines starken Begriffs von Inkommensurabilität, dann wird sofort erkennbar, dass hier beide Problemtypen, die mit der Inkommensurabilität verbunden sind, vorliegen. Sowohl Probleme der Verständigung wie auch Probleme der epistemischen Rechtfertigung sind zu beobachten. Wie äussern sich diese beiden Problemtypen konkret?

Greifen wir zur Veranschaulichung auf eine Beispielsituation zurück. Nehmen wir an, dass die EOT Ergebnisse des institutionsökonomischen Ansatzes integrieren möchte. Exemplarisch können folgende *Probleme der Verständigung* genannt werden. Die Aussage des institutionsökonomischen Ansatzes, dass bei niedriger Spezifität der Transaktion die Koordinationsform des Marktes für die Leistungserstellung genutzt werden soll (vgl. Picot et al., 2008, S. 70), lässt sich nicht in das Sprachspiel der EOT übersetzen, wenn man eine starke Ausprägung der Inkommensurabilitätsrelation voraussetzt. Die Begriffskategorien funktionieren für

[21] Der beschränkte Umfang der Untersuchung verunmöglicht eine vertiefte Diskussion dieser Thematik. Siehe hierzu weiterführend Winter (1999, S. 148 ff.), von Glaserfeld (1985, 1998, S. 186 ff.), von Foerster (1993, S. 25 ff.) sowie von Foerster und Pörksen (2003, S. 25 ff., 59 ff.).

[22] Siehe erneut Abbildung 7

[23] Der Leser kann an dieser Stelle einwenden, dass die intertheoretischen Implikationen einen objekttheoretischen Zugang zur Inkommensurabilitätsthematik voraussetzen. Sobald die Beziehung der EOT zu konkurrierenden organisationstheoretischen Ansätzen ins Zentrum rückt, verlassen wir auf den ersten Blick die metatheoretische Perspektive. Dies ist jedoch nicht der Fall, da die EOT als Rahmentheorie angelegt ist und damit konkurrierende Ansätze in das eigene Theoriegebäude einverleibt. Es handelt sich also immer noch um eine theorieinterne und somit metatheoretische Perspektive.

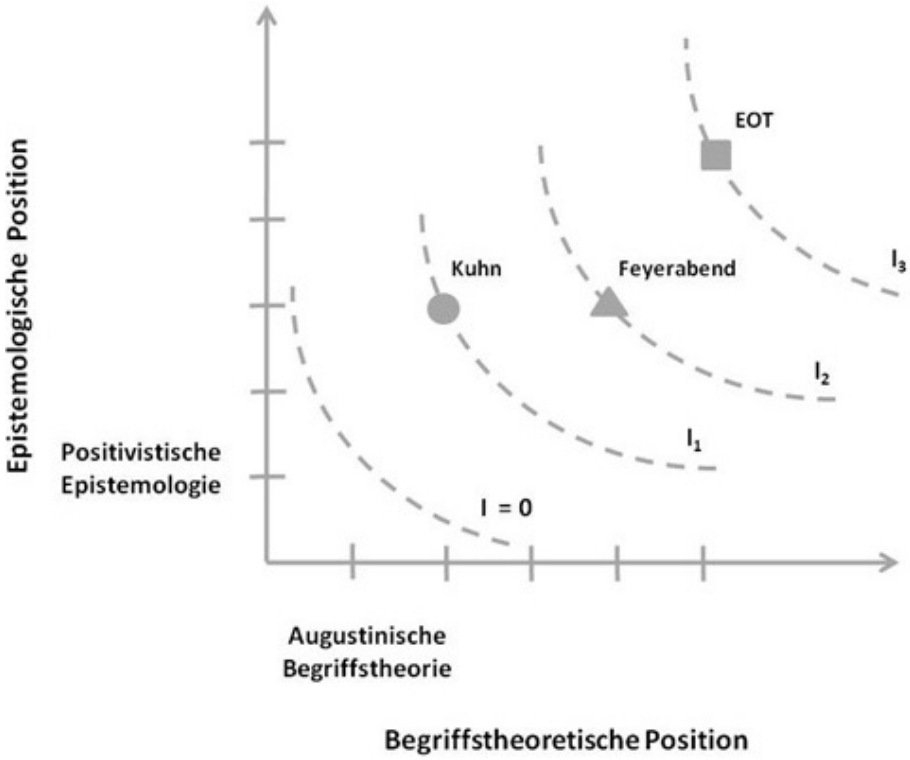

Abbildung 8: Positionierung des wissenschaftstheoretischen Fundaments der EOT (Quelle: eigene Darstellung).

die beiden Aussagensysteme unterschiedlich, so dass keine adäquate Übersetzung möglich ist. Des Weiteren stellt sich die berechtigte Frage, wie bspw. diese nicht übersetzbare Aussage sinnvollerweise auf das Aussagensystem der EOT bezogen werden kann. Auch wenn sich innerhalb der EOT Aussagen zur optimalen Ausgestaltung von Leistungserstellungsprozessen findet (vgl. exemplarisch Kirsch, 1998, S. 106; Kirsch und Guggemos, 1999b, S. 38, 74), dann ist nicht klar, wie diese Aussagen auf die erwähnte These des institutionsökonomischen Ansatzes zu beziehen sind. Obwohl in beiden organisationstheoretischen Ansätzen dieselben Begriffe verwendet (bspw. „Transaktion“ und „Leistungserstellung“) und analoge Sachverhalte diskutiert werden (bspw. Lieferantennetzwerke), ist aufgrund der Eigenständigkeit der Sprachspiele eine Bezugnahme zwischen den Aussagen nicht möglich.

Wechseln wir auf die Seite der *Probleme der epistemischen Rechtfertigung*, so zeigt sich in der Beispielssituation ein ähnliches Bild. Der institutionsökonomische Ansatz propagiert eine kausallogisch motivierte, empirische Forschungsmethodik (vgl. Picot et al., 2008, S. 45 f.), während sich bei der EOT ein klares Bekenntnis zu einem systemisch-explorativen und damit nicht-empirischen Forschungsansatz findet (vgl. Kirsch et al., 2007, S. 43). Darüber hinaus ist auch eine Diskrepanz zwischen den beiden Gegenstandsbereichen dieser organisationstheoretischen Ansätze zu beobachten. Während die EOT Organisationen als evolvierende soziale Phänomene betrachtet, die mit einer radikal offenen Zukunft konfrontiert sind, konzeptualisiert der institutionsökonomische Ansatz sein Forschungsobjekt über weite Teile als Bündel von Transaktionen (vgl. Picot et al., 2008, S. 57). Unter der Voraussetzung einer starken Inkommensurabilitätsrelation ist nicht von einer Übereinstimmung des Forschungsobjektes auszugehen.

Mit der Exemplifikation der beiden Problemtypen wird erkennbar, dass die radikal pluralistische Orientierung der EOT – in Form des Beziehungsgeflechts mit konkurrierenden organisationstheoretischen Ansätzen – erhebliche Probleme generiert und damit als wesentlicher Pfeiler der Theorie (vgl. Kirsch, 1998, S. 129 ff.) nicht taugt.

Soweit die Ausführungen zur ersten intertheoretischen Implikation. Die Zweite lässt sich wie folgt formulieren:

> *Befund 2:* Die Konzeption der EOT als Theorieprozess, der sich als einheitliches Ganzes aus Momentaufnahmen der EOT zu verschiedenen Zeitpunkten zusammensetzt, ist problematisch.

Betrachtet man die EOT zu unterschiedlichen Zeitpunkten ihres Entwicklungsstadiums, dann kann man einen Vergleich zwischen diesen jeweils statischen Aussagensystemen anstreben.[24]

Wie die Ausführungen zum ersten Befund deutlich gemacht haben, artikulieren sich in derartigen Situationen von

[24] Diese Betrachtung ist wiederum eine intertheoretische. Obwohl hier keine Relation zwischen der EOT und konkurrierenden organisationstheoretischen Ansätzen besteht, ist eine relationale Beziehung zwischen zwei Aussagensystemen – die EOT in zwei Ausprägungen – zu konstatieren.

Theorievergleichen Probleme der Verständigung und der epistemischen Rechtfertigung. Dieselben Probleme sind insbesondere auch dann beobachtbar, wenn man die EOT zu unterschiedlichen Zeitpunkten ihrer Entwicklung betrachtet. Indem die EOT von einer radikalen Offenheit der organisationalen Zukunft ausgeht, schreibt sie ihrem Gegenstandsbereich eine Evolutionsdynamik zu, die mit den Ressourcen des jeweils gegenwärtigen organisationstheoretischen Sprachspiels nicht erfassbar ist.

Damit geht die EOT jedoch indirekt davon aus, dass zeitlich spätere, evolvierte Ausprägungen der Theorieanlage einen radikal anderen Inhalt umschliessen können, als dies bei dem heutigen Stand der EOT der Fall ist. Stellen wir uns bspw. den Fall vor, dass sich die Gruppe relevanter externer Akteure einer Organisation, welche die EOT voraussetzt, erheblich ändert. So könnte die EOT in einer zukünftigen Ausprägung Web 2.0 basierte Plattformen als relevante Akteure für eine Organisation akzeptieren. Mit der Integration dieses Beobachtertyps ändern sich das Gefüge der organisationalen Akteure sowie zentrale Einsichten der EOT zu Verhaltensmustern, Kommunikationsströmen und Interferenzen zwischen den unterschiedlichen Konstruktionen organisationaler Realität. Versucht man diese neuen Einsichten auf die älteren Befunde der EOT zu beziehen, dann liegen Übersetzungs- und Artikulationsprobleme auf der Hand. Zugleich ist eine Veränderung des Forschungsgegenstandes zu konstatieren, der zu einer Änderung der propagierten Forschungsmethodik führt.

Nimmt man diesen Befund ernst, dann erscheint das Selbstverständnis der EOT als evolvierendes Theoriegebäude problematisch. Vor dem Hintergrund einer starken Ausprägung der Inkommensurabilitätsrelation ist nicht klar, wie die unterschiedlichen Ausprägungen der EOT sinnvoll aufeinander bezogen werden können und wie damit eine einheitliche Theoriekonzeption entstehen kann. Mit diesen zwei kritischen Befunden schliessen wir die Diskussion der intertheoretischen Implikationen der Inkommensurabilitätsthematik ab. Wenden wir uns den intratheoretischen Implikationen zu.

3.3.3. Intratheoretische Implikationen der Inkommensurabilität

Wir haben die intratheoretischen Implikationen anhand der Frage charakterisiert, was mit den Aussagen der EOT über ihren Gegenstandsbereich passiert, wenn wir die Inkommensurabilitätsthematik ernst nehmen. Im Fokus stehen die Thesen der EOT zu organisationalen Akteuren und Problemen, zu organisationalen Entwicklungspfaden sowie zu polyzentrischen Unternehmensstrukturen. Nachfolgend werden im Hinblick auf diese drei Themenfelder jeweils zwei kritische Befunde identifiziert und erläutert.

Beginnen wir die Diskussion mit dem ersten kritischen Befund zum *Akteurs- und Problemverständnis der EOT*:

> *Befund 3:* Die Vorstellung von einer problemlosen Wissenszirkulation zwischen den organisational relevanten Akteuren und einem verstehenden Zugang derselben untereinander ist nicht haltbar.

Wir haben gesehen, dass die EOT im Umfeld einer Organisation Akteure verortet, die als Beobachter organisationaler Phänomene fungieren. Diese Beobachter generieren jeweils aufgrund ihrer unterschiedlichen lebensweltlichen Kontexte divergente Wirklichkeitskonstruktionen. Zwischen diesen Akteursgruppen untereinander und der Organisation im Zentrum konstatiert die EOT wechselseitige Beobachtungen und Interaktionen (vgl. Kirsch, 1998, S. 46; Kirsch und Weber, 1999d, S. 9, 114 f.). Im Kontext unseres Bezugsrahmens ist dieses Beziehungsgeflecht als eine Menge von zweigliedrigen Relationen zwischen Aussagensystemen zu begreifen. Unter der Voraussetzung einer starken Ausprägung der Inkommensurabilitätsrelation treten hier insbesondere Probleme der Verständigung auf.

Heben wir zum Zweck der Konkretisierung dieses Punktes einen Akteurstyp hervor: Berater, die sich als externe Experten[25] mit organisationalen Problemen befassen. Zwischen Beratungsunternehmen und den übrigen Akteursgruppen besteht gemäss der EOT ein dichtes Beziehungsnetz (vgl. Kirsch, 1998, S. 20, 54 f.; Kirsch und Guggemos, 1999b, S. 106). Reports von Finanzanalysten, neuste betriebswissenschaftliche Erkenntnisse sowie organisationsinterne Dokumente (bspw. ausgewählte Strategiepapiere oder Marktstudien) finden als Datenmaterial Eingang in den Beratungsprozess. Dieser Informationsfluss ist nicht nur einseitig ausgestaltet. Ergebnisse von Beratungsprojekten diffundieren in den organisationalen Alltag und werden von organisationstheoretischen Ansätzen aufgegriffen.[26] Nimmt man die These der EOT ernst, dass die vier Akteurstypen Berater, Wissenschaftler, Analysten und organisationsinterne Akteure jeweils in einem unterschiedlichen lebensweltlichen Kontext agieren, dann zeigen sich zwischen den jeweiligen Wirklichkeitskonstruktionen sowohl Übersetzungs- wie auch Artikulationsprobleme. Exemplarisch ist bspw. ein Übersetzungsproblem vorhanden, falls der Versuch unternommen wird, die Lösungspräsentation einer Beratungsfirma zur Frage nach möglichen Prozessverbesserungen in ein organisationsinternes Dokument zu übersetzen. Ein Artikulationsproblem liegt bspw. dann vor, wenn sich die organisationsinternen Akteure fragen, wie sie Beobachtungen der Beratungsfirma sinnvollerweise auf organisationsinterne Dokumente und Abläufe beziehen sollen.

Das dichte Beziehungsnetz zwischen den organisationalen Akteuren, welches die EOT skizziert, ist demzufolge problembehaftet. Damit ist dieses Element nicht geeignet, die

[25] Hier findet eine Verengung des Beratungsmarktes statt. Neben der „klassischen“ Beratung, die im Vermitteln von Expertenwissen besteht, existiert auch eine systemische Unternehmensberatung. Die Verengung ist dem begrenzten Umfang der Untersuchung geschuldet. Siehe zur systemischen Beratung weiterführend Nagel (2001, S. 14 ff.) und Winter (1999, S. 199 ff.)

[26] Als exemplarisches Beispiel hierfür kann die BCG Matrix dienen, die es erlaubt, anhand der zwei Dimensionen Marktwachstum und relative Marktgrösse Portfolioentscheide betreffend unternehmerischen Geschäftsfeldern zu fundieren (vgl. Thommen und Achleiter, 2006, S. 941 ff.).

pluralistische Intention einer Eingliederung aller relevanten Akteure und deren Beziehungen untereinander in die EOT umzusetzen.

Wenden wir uns dem zweiten kritischen Befund zum Akteurs- und Problemverständnis der EOT zu:

> *Befund 4:* Die Strategie der echten Komplexitätsbejahung zur Lösung organisationaler Probleme, bei der mehrere organisationale Akteure beteiligt sind, ist nicht zielführend.

Indem die EOT die Vielschichtigkeit organisationaler Probleme, die aufgrund der beteiligten Akteure entsteht, betont, taxiert sie Probleme folgerichtig als Multi-Kontext-Probleme. Die zielführende Strategie zur Lösung solcher Probleme besteht gemäss der EOT in einer echten Komplexitätsbejahung, welche die Vielzahl an divergenten Wirklichkeitskonstruktionen der betroffenen Akteure ernst nimmt. Die Beziehung zwischen den Problembeschreibungen und Lösungsartikulationen, welche die Akteure entwickeln, die an einem organisationalen Problem beteiligt sind, lässt sich wiederum als zweigliedrige Relation zwischen Aussagensystemen begreifen. Analog zu Befund drei treten unter der Annahme einer starken Ausprägung der Inkommensurabilitätsrelation insbesondere Probleme der Verständigung auf.

Stellen wir uns zur Veranschaulichung dieses Punktes ein konkretes organisationales Problem vor. Das Pharmaunternehmen X hat einen Wirkstoffproduzenten Y im Rahmen einer echten Fusion übernommen und sieht sich im Zuge der Integrationsbemühungen mit dem Problem konfrontiert, wie sie die unternehmensweit unterschiedlichen Arbeitszeitenregelungen vereinheitlichen soll. Von diesem Problem ist nicht nur unmittelbar die Belegschaft der entstandenen Organisation A betroffen. In Bezug auf diese arbeitsrechtliche Fragestellung spielen externe Akteure ebenfalls eine zentrale Rolle. Insbesondere sind hier staatliche Regulierungsinstanzen, Gewerkschaften, weitere Interessenverbände, mediale Beobachter wie auch Teile der wissenschaftlichen Gemeinschaft (bspw. ein auf Fragen des Arbeitsrechts spezialisiertes rechtswissenschaftliches Institut) zu nennen. All diese Akteure operieren gemäss der EOT als eigenständige Beobachter, die vor dem Hintergrund ihres unterschiedlichen lebensweltlichen Kontexts als Wirklichkeitskonstrukteure auftreten und damit jeweils individuelle Einschätzungen der arbeitsrechtlichen Problemsituation anfertigen.

Nehmen wir die Aufforderung der EOT ernst und realisieren in unserem Fallbeispiel die Problemlösungsstrategie der echten Komplexitätsbejahung. Konkret bedeutet dies, dass wir die Situations- und Problembeschreibungen einer grösstmöglichen Zahl der organisational relevanten Akteure bei der Behandlung der Frage nach der Ausgestaltung der Arbeitszeitenregelung einbeziehen müssen. Indem wir zulassen, dass die Akteure untereinander in eine selbstorganisierende Interaktion treten, im Zuge derer sie die relevanten Themen und relevanten Akteure definieren, tragen wir der Pluralität der Wirklichkeitskonstruktionen adäquat Rechnung (vgl. Kirsch et al., 2007, S. 87). An diesem Punkt der Argumentation lassen sich die Probleme der Verständigung lokalisieren, die sich für die EOT im Zuge der impliziten Bejahung einer starken Ausprägung der Inkommensurabilitätsrelation ergeben. Es ist keinesfalls klar, wie die unterschiedlichen Problem-, Situations- und Lösungsbeschreibungen der Akteure sinnvoll ineinander übersetzt und aufeinander bezogen werden können. Die Pluralität an lebensweltlichen Kontexten verhindert gerade die Fruchtbarkeit der Interaktion, welche die EOT als Idealbild skizziert.

Die Veranschaulichung anhand des arbeitsrechtlichen Fallbeispiels hat eines deutlich gemacht: Die Strategie der echten Komplexitätsbejahung ist aufgrund des wissenschaftstheoretischen Fundaments der EOT keine zielführende Strategie, um organisationale Probleme zu konzeptualisieren und zu lösen. Auch dieses Element der EOT ist damit wenig geeignet, organisationale Zusammenhänge pluralistisch zu konzeptualisieren.

Als nächstes ist das *Bild organisationaler Entwicklung*, welches die EOT skizziert, zu behandeln. Wiederum können zwei kritische Befunde identifiziert werden. Beginnen wir mit dem Konstrukt des Sinnmodells.

> *Befund 5:* Das Konstrukt des Sinnmodells, das als gemeinsam geteilter lebensweltlicher Hintergrund der organisationsinternen Akteure prägend für Problemdefinitionen, Situationsbeschreibungen und Lösungsentwicklungen ist, ist nicht haltbar.

Die Ausführungen zu den zentralen Elementen der EOT haben deutlich gemacht, dass der Begriff des Sinnmodells jene Inhalte umfasst, die handlungsanleitenden Charakter in Bezug auf Problemdefinitionen, Situationsbeschreibungen und Lösungsentwicklungen innerhalb des organisationalen Alltags haben. Die Handlungsanleitung ist nicht im Sinne von konkreten Weisungen zu verstehen, sondern als generelle Orientierungshilfe für das Handeln im organisationalen Kontext (vgl. Kirsch et al., 1999a, S. 133; Kirsch und Heeckt, 2001, S. 244). Das Sinnmodell ist ein Konstrukt, das den Zweck einer Organisation festlegt und sich somit einer Organisation als Ganzes zuschreiben lässt. So verstanden bejaht das Sinnmodell die Frage, ob von einer organisationalen Lebenswelt – im Sinne eines Kontexts, der für alle organisationsinternen Akteure prägenden Charakter hat – gesprochen werden kann (vgl. Kirsch et al., 1999a, S. 123).[27]

Betrachten wir das Konstrukt des Sinnmodells vor dem Hintergrund der Inkommensurabilitätsrelation. Nimmt man die These der EOT ernst, dass sich innerhalb einer Organisation Interaktionen zwischen Akteuren abspielen, die jeweils unterschiedlichen lebensweltlichen Kontexten verhaftet sind

[27] Diese Formulierung erweckt den Eindruck einer Gleichsetzung der Begriffe „Sinnmodell“ und „Lebenswelt“. Dies wäre unzulässig, da die Autoren der EOT dezidiert für eine saubere Trennung dieser Begriffe argumentieren (vgl. insbesondere Kirsch et al., 1999a, S. 121 ff.). Zum Zwecke unserer Argumentation interpretieren wir das Sinnmodell als organisationalen Kontext, von dem aus die Akteure Situationen beschreiben und Probleme analysieren. So betrachtet ist das Sinnmodell ein lebensweltlicher Kontext, da es prägend für die Wirklichkeitskonstruktionen und Handlungen der organisationalen Akteure ist.

(vgl. ebd., S. 260), dann lässt sich dieses Konstrukt problematisieren. Die wechselseitigen Beziehungen zwischen den organisationsinternen Akteuren sind als relationale Beziehungen zwischen Aussagensystemen modellierbar.[28] Geht man von einer starken Ausprägung der Inkommensurabilitätsrelation aus, dann fallen wiederum die Probleme der Verständigung zwischen zwei inkommensurablen Aussagensystemen ins Auge. Unterschiedliche Akteure können mit Übersetzungs- und Artikulationsproblemen konfrontiert sein, wenn sie versuchen mit anderen Akteuren zu interagieren. Das naheliegende Beispiel für eine derartige Konstellation ist die Situation, bei der zwei Mitarbeiter mit unterschiedlichem kulturellem Hintergrund bei einem Projekt zusammenarbeiten. So kann bspw. bei einer internationalen Unternehmensberatung aufgrund des Staffings ein US-Amerikaner gemeinsam mit einem Chinesen auf einem Projekt arbeiten. Das Stichwort „Übersetzungs- und Artikulationsproblem“ meint in diesem Fall jedoch nicht die (offensichtliche) Sprachbarriere zwischen der englischen und chinesischen Sprache. Vielmehr kann der unterschiedliche lebensweltliche Kontext unter der Voraussetzung einer starken Ausprägung der Inkommensurabilitätsrelation dazu führen, dass die zwei Mitarbeiter trotz gemeinsam geteiltem englischen Wortschatz nicht in der Lage sind, die Aussagen des Gegenübers sinnvoll zu interpretieren.

Nehmen wir diese Erkenntnis aus dem Beispiel ernst, dann wird Folgendes klar: Unter Berücksichtigung der Vielzahl an lebensweltlichen Kontexten, die für organisationale Interaktionen relevant sind, ist nicht ersichtlich, wie ein Konstrukt entstehen kann, dass organisationsweit – und damit über die divergenten lebensweltlichen Kontexte hinweg – handlungsanleitenden Charakter entwickeln kann. Das Sinnmodell ist ein gemeinsam geteilter Kontext, der für Problemdefinitionen, Situationsbeschreibungen und Lösungsentwicklung zumindest paradigmatisch handlungsanleitend ist. Diese Möglichkeit der Erzeugung eines gemeinsamen Kontexts – und damit die Redeweise von einem organisationsweiten Sinnmodell – ist nicht plausibel, sobald man die Pluralität der organisational relevanten Lebenswelten und die Ausprägung der Inkommensurabilitätsrelation adäquat berücksichtigt.

Die Folgen dieser Problematik werden sichtbar, wenn wir den zweiten kritischen Befund mit in die Argumentationskette aufnehmen.

> *Befund 6:* Organisationale Entwicklung, organisationales Lernen und organisationale Wertsteigerung lassen sich nicht über das Konstrukt des Sinnmodells definieren.

Die EOT versteht organisationale Entwicklung als Übergang zwischen den zentralen Sinnmodellen Instrumentalmodell, Überlebens- resp. Bestandsmodell und Fortschrittsmodell. Dass diese Charakterisierung von organisationaler Entwicklung problematisch ist, folgt direkt aus den vorangegangen Bemerkungen zum Konstrukt des Sinnmodells.

Akzeptiert man den Befund, dass das Konstrukt des Sinnmodells vor dem Hintergrund einer starken Ausprägung der Inkommensurabilitätsrelation nicht haltbar ist, dann verschwindet der Massstab anhand dessen bestimmt werden kann, ob eine organisationale Entwicklung stattfindet oder nicht. Denn die unterschiedlichen Sinnmodelle und die Übergänge zwischen diesen bilden den Referenzpunkt, um organisationale Entwicklung definieren zu können. Das Wegfallen des Referenzpunktes hat jedoch nicht nur Folgen für das Verständnis von organisationaler Entwicklung.

Zum einen findet sich in der EOT die These, dass organisationales Lernen ebenfalls über das Konstrukt des Sinnmodells definiert werden kann. Eine Organisation kann dann einen Lernerfolg vorweisen, wenn sie im Hinblick auf die Offenheit ihrer eigenen Zukunft, die Handlungsvorgaben an ein höheres Sinnmodell koppelt (vgl. Kirsch et al., 1999a, S. 338). Analog entfällt dieser Zugang zu organisationalem Lernen, wenn man das Konstrukt des Sinnmodells zurückweist. Zum anderen legt die EOT begrifflich fest, dass organisationale Wertsteigerung genau dann vorliegt, wenn sich eine Organisation auf der Stufe eines Sinnmodells in Bezug auf die jeweils wesentliche Leistungsdimension signifikant verbessert (vgl. Kirsch und Heeckt, 2001, S. 246). Dies ist bspw. dann der Fall, wenn auf der Ebene des Instrumentalmodells die Eigenkapitalrentabilität über ein Geschäftsjahr signifikant gesteigert wird (vgl. ebd.). Es liegt auf der Hand, dass auch in Bezug auf diese begriffliche Festlegung der EOT das analoge Problem auftritt, wenn der Referenzpunkt „Sinnmodell“ entfällt.

Zusammenfassend lässt sich konstatieren, dass durch die Problematisierung des Konzepts des Sinnmodells ein Baustein entfällt, der eine wichtige explanatorische Funktion für die EOT erfüllte.

Abschliessend ist die *Behandlung polyzentrischer Unternehmensstrukturen* durch die EOT genauer zu betrachten. Ein erster Befund lässt sich wie folgt formulieren:

> *Befund 7:* Die These von einem Polyzentrismus im schwachen Sinne, der auf der Möglichkeit der Homogenisierung divergenten Lebenswelten besteht, ist nicht haltbar.

Die EOT erhebt den Anspruch durch die Bereitstellung eines Sprachspiels, mit welchem sich der Pluralismus an Lebenswelten begrifflich fassen lässt, für die theoretische Diskussion polyzentrischer Unternehmensstrukturen prädestiniert zu sein (vgl. Kirsch et al., 2009, S. 292; Kirsch, 1996, S. 252). In den Ausführungen zu den zentralen Elementen der EOT konstatierten wir, dass die EOT immer dann von einer polyzentrischen Unternehmensstruktur spricht, wenn ein Unternehmen die Geschäftstätigkeit internationalisiert oder Unternehmensverbindungen eingeht. Innerhalb einer derartigen polyzentrischen Unternehmensstruktur finden sich des Weiteren Partialsysteme, in denen sich jeweils ein eigener lebensweltlicher Kontext ausbildet.

[28] Diese Charakterisierung mag auf den ersten Blick irritieren. Führt man sich jedoch vor Augen, wie viel der organisationsinternen Interaktion (insbesondere in Dienstleistungsunternehmen) sich über Email- und Telefonverkehr abspielt, dann wird die Modellierung über Aussagensysteme einsichtig.

Betrachten wir die Konstellation der Partialsysteme etwas genauer. Die Verhältnisse zwischen den Partialsystemen in einer polyzentrischen Unternehmensstruktur können wiederum als zweigliedrige, relationale Beziehungen zwischen Aussagensystemen interpretiert werden. Vor dem Hintergrund unseres metatheoretischen Bezugsrahmens und dem wissenschaftstheoretischen Fundament der EOT ist zwischen diesen Partialsystemen eine Inkommensurabilitätsrelation mit starker Ausprägung zu konstatieren. In Folge dieser Inkommensurabilitätsrelation sind besonders Probleme der Verständigung virulent. Dieser Punkt lässt sich wiederum anhand eines kurzen Beispiels exemplifizieren.

Betrachten wir den Lebensmittelkonzern Z, der aus zwei Tochtergesellschaften besteht. Die Gesellschaft V ist in Südostasien domiziliert und für den Einkauf und die primäre Verarbeitung von Rohstoffen zuständig. Die Gesellschaft W in Süddeutschland empfängt die aufbereiteten Rohstoffe und verarbeitet diese zu Lebensmitteln. Wir attestieren im Sinne der EOT, dass der Konzern Z eine polyzentrische Unternehmensstruktur aufweist und sich innerhalb der Tochtergesellschaften V und W Partialsysteme ausgebildet haben, die als handlungsanleitenden Kontext für die jeweiligen organisationalen Tätigkeiten der Tochtergesellschaften prägenden Charakter besitzen. Akzeptieren wir die These einer starken Inkommensurabilitätsrelation zwischen diesen beiden Partialsystemen, dann hat dies folgende exemplarische Konsequenzen. Die Übersetzung von Berichten, Einschätzungen und Anweisungen der Tochtergesellschaft V über Logistikorganisation, Rohstoffmärkte und lokale politische Gegebenheiten in Papiere der Tochtergesellschaft W ist problematisch. Zudem tritt das Problem auf, dass sich die Berichte, Einschätzungen und Anweisungen der Tochtergesellschaft V nicht ohne Weiteres auf die Einschätzungen und organisationsinternen Dokumente von W beziehen lassen. In diesem Sinn sind erhebliche Verständigungsschwierigkeiten zwischen den beiden Partialsystemen festzustellen.

Welche Konsequenzen können aus dieser Einsicht abgeleitet werden? In Bezug auf die Unterscheidung der EOT zwischen einem Polyzentrismus im starken und schwachen Sinnen ergibt sich eine direkte Folge. Diese Unterscheidung beruht auf der Annahme, dass Partialsysteme einen handlungsanleitenden und damit homogenisierenden Kontext teilen können (vgl. Kirsch, 1996, S. 274). Dieser Kontext entsteht im Zuge der gemeinsamen Interaktion zwischen den beiden Partialzentren (vgl. Kirsch, 1998, S. 100 f.). Das Vorliegen eines derartigen Kontexts ist das definierende Merkmal für einen Polyzentrismus im schwachen Sinne. Nimmt man die vorangehende Argumentationslinie ernst, dann ist die Möglichkeit eines derartigen homogenisierenden Kontexts zwischen zwei Partialsystemen aufgrund der starken Ausprägung der Inkommensurabilitätsrelation ausgeschlossen. Die unterschiedlichen lebensweltlichen Kontexte im Zusammenhang mit den auftretenden Problemen der Verständigung verunmöglichen die Erzeugung eines gemeinsam geteilten, handlungsanleitenden Vorrats an Wissensbeständen zwischen den beiden Partialsystemen.

Zusammenfassend können wir dementsprechend Folgendes festhalten: Falls die EOT konsistent aus ihrem wissenschaftstheoretischen Fundament entwickelt wird, kann sie in Bezug auf polyzentrische Unternehmensstrukturen nur einen Polyzentrismus im starken Sinne propagieren.

Ausgehend von der Zurückweisung der Vorstellung eines Polyzentrismus im schwachen Sinne, können wir eine weitere Implikation ableiten:

> *Befund 8:* Die Möglichkeit strategischer Führung im Kontext von polyzentrischen Unternehmensstrukturen muss relativiert werden.

Die EOT konstatiert, dass innerhalb von polyzentrischen Unternehmensstrukturen strategische Führung wirksam werden kann (vgl. Kirsch, 1996, S. 289). Unter strategische Führungsaktivitäten fallen all jene Managementaktivitäten, die orientierend, strukturzentriert und ganzheitlich die Handlungen innerhalb einer Organisation leiten sollen (vgl. ebd., S. 288 f.).[29] Strategische Führung ist auch dann möglich, wenn den Partialzentren ein Eigensinn in Form eines eigenständigen lebensweltlichen Kontexts zugesprochen wird (vgl. ebd.). Innerhalb der EOT findet sich jedoch zugleich eine Einschränkung der Möglichkeit strategischer Führung in polyzentrischen Unternehmensstrukturen. Sobald ein Polyzentrismus im starken Sinne vorliegt, ist das Auftreten und die erfolgreiche Umsetzung zentraler strategischer Führungsvorgaben (wie bspw. ein konzernweites Produkt-Controlling) eher unwahrscheinlich (vgl. ebd., S. 275). Im Gegensatz dazu, stuft die EOT strategische Führungsvorhaben im Kontext eines Polyzentrismus im schwachen Sinne als realisierbares Unterfangen ein. Führen wir diesen Befund mit unserer Einsicht zusammen, dass nur sinnvoll von einem Polyzentrismus im starken Sinne gesprochen werden kann, dann lässt sich Folgendes festhalten: Die EOT kann in Bezug auf polyzentrische Unternehmensstrukturen nicht behaupten, dass das Auftreten von strategischer Führung grundsätzlich wahrscheinlich ist. Damit erfährt die These der strategischen Führung innerhalb derartiger Unternehmensstrukturen eine weitere Relativierung.

Mit diesem letzten Befund schliessen wir die Diskussion der Implikationen aus der Inkommensurabilitätsthematik für die EOT ab. Die Eruierung der Ausprägung der Inkommensurabilitätsrelation innerhalb der EOT sowie die Identifizierung von acht kritischen inter- und intratheoretischen Befunde haben zum einen die Funktionsweise unseres metatheoretischen Bezugsrahmens verdeutlicht. Zum anderen liessen sich weitreichende Implikationen für die EOT ableiten und an konkreten, organisationalen Beispielen exemplifizieren. Damit ist auch die zweite Forschungsfrage beantwortet.

4. Schlussbemerkungen

Ausgangspunkt der Untersuchung bildete der offenkundige Theoriepluralismus im Feld der Organisationstheorie. Die

[29] Die Behandlung des Phänomens der (strategischen) Führung in der EOT wäre an dieser Stelle weiter auszuführen. Aufgrund des beschränkten Umfangs dieser Untersuchung finden nur diejenigen Aspekte Erwähnung, die für die vorgetragene Argumentationslinie notwendig sind.

Analyse der Diskussion dieses Phänomens in der organisationstheoretischen Literatur machte deutlich, dass der Begriff der Inkommensurabilität den Dreh- und Angelpunkt der Debatte um die Vor- und Nachteile dieses Theoriepluralismus markiert. Zugleich stellten wir fest, dass sich in der organisationstheoretischen Literatur keine vertiefte Auseinandersetzung mit den Grundlagen der Inkommensurabilitätsthese finden lässt. Das Ziel der Untersuchung bestand darin, diese Forschungslücke zu schließen, indem ein solides begriffliches Fundament generiert und dessen Implikationen exemplarisch aufgezeigt werden.

In einem ersten Schritt wurde ein metatheoretischer Bezugsrahmen zur Inkommensurabilitätsthematik hergeleitet. Ausgehend von vier zentralen Elementen des Kuhnschen Begriffs von Inkommensurabilität wurden zwei – einzeln notwendige und zusammen hinreichende – Bedingungen entwickelt, die das Auftreten der Inkommensurabilitätsrelation ermöglichen. Die Bedingungen „nicht-augustinische Begriffstheorie“ und „nicht-positivistische Epistemologie“ etablieren ein graduelles Verständnis von Inkommensurabilität. Sie greifen dabei auf einer metatheoretischen Ebene, da sie die Beurteilung der Ausprägung der Inkommensurabilitätsrelation für eine bestimmte Theorieanlage ermöglichen. Diese abstrakten Einsichten bildeten das Fundament, um in einem zweiten Schritt die Implikationen der Inkommensurabilitätsthematik auszuarbeiten. Als geeignetes Transferobjekt identifizierten wir die EOT. Vor dem Hintergrund der Skizze fünf zentraler Elemente dieses organisationstheoretischen Ansatzes wurde die Ausprägung der Inkommensurabilitätsrelation sowie deren inter- und intratheoretischen Implikationen für die EOT sichtbar gemacht. Die Implikationen kondensierten in acht kritischen Befunden, die in pointierter Form die weitreichenden Konsequenzen einer adäquaten Behandlung der Inkommensurabilitätsthematik innerhalb der EOT aufzeigten.

Treten wir einen Schritt zurück und ordnen diese Einsichten in den Diskussionszusammenhang ein, der Anlass zur Untersuchung gab. Prüfstein für die Frage, ob der Theoriepluralismus ein positiver und damit anzustrebender Zustand für die Organisationstheorie darstellt, ist die Inkommensurabilitätsthematik. Ist die These inkommensurabler Perspektiven zwischen unterschiedlichen organisationstheoretischen Ansätzen haltbar, dann haben die vorgetragenen Erläuterungen aufgezeigt, dass eine fruchtbare Integration organisationstheoretischer Forschungsbemühungen nicht realistisch ist. Die Vorstellung, dass sich durch die Vielzahl an Theorien ein mosaikartiger und umfassender Zugang zum Phänomen „Organisation“ eröffnet, bleibt in diesem Fall eine Fiktion und theoretischen Assimilationsbemühungen wären ein grösseres Gewicht einzuräumen.

Die zentrale Frage lautet dementsprechend: Ist die These inkommensurabler Perspektiven im Bereich der Organisationstheorie haltbar? Ein fruchtbarer Weg diese Fragestellung anzugehen, findet sich unseres Erachtens in der vorliegenden Untersuchung. Erst wenn man sich über die begriffstheoretischen und epistemologischen Zugeständnisse, die mit der Inkommensurabilitätsrelation verbunden sind, im Klaren ist, kann man eruieren, ob man vor dem Hintergrund der eigenen wissenschaftstheoretischen Annahmen tatsächlich bereit ist eine derartige These zu vertreten. Denn im Zuge der Argumentation wurde deutlich, dass die begriffstheoretischen und epistemologischen Zugeständnisse im Kontext der Inkommensurabilitätsthematik von erheblicher Natur sind. Es ist dieser metatheoretische Zugang, den wir als zielführende Strategie vorschlagen. Nur ein solcher berücksichtigt die Komplexität der Inkommensurabilitätsthematik adäquat und entzieht sich einer vorschnellen, objekttheoretischen Untersuchung, bei der organisationstheoretische Ansätze zueinander in Beziehung gesetzt und auf Inkommensurabilitätsverhältnisse hin untersucht werden. Kurzum: Der Theoriepluralismus im Feld der Organisationstheorie ist erst dann beurteilbar, wenn man sich im Klaren darüber ist, welche begriffstheoretischen und epistemologischen Zugeständnisse man bereit ist einzugehen. Dadurch wird zugleich der Raum für eine spezifische organisationstheoretische Forschungsbemühung eröffnet, auf die wir abschliessend zu sprechen kommen. Diese Einschätzung des Theoriepluralismus basiert auf den Voraussetzungen unseres Gedankengangs. Eine kritische Reflexion der vorgetragenen Argumentationskette ist demzufolge angezeigt. Insbesondere sind zwei Beschränkungen zu nennen.

Erstens ist die systematische Perspektive auf die Inkommensurabilitätsthematik zu reflektieren. Durch diese Perspektivenwahl haben wir uns gegen eine rein exegetische Auseinandersetzung mit den Schriften von Kuhn und Feyerabend ausgesprochen. Dieser – für unsere Zwecke fruchtbare – Zugang ist mit zwei Problemen verbunden. Zum einen verlieren wir mit der systematischen Betrachtung die exegetische Bodenhaftung. Damit ist der Umstand bezeichnet, dass im Zuge der Argumentation Aspekte der Inkommensurabilitätsthematik ausgeblendet wurden, die bei Kuhn sowie Feyerabend von zentraler Natur sind. Das Faktum, dass der Begriff der Inkommensurabilität bei Feyerabend und Kuhn für den Bereich der Naturwissenschaften reserviert wurde (vgl. Hoyningen-Huene, 1989, S. 16; Oberheim und Hoyningen-Huene, 2009, S. 10), floss bspw. nicht in unsere Reflexion mit ein. Damit erscheint der Transfer dieser Begrifflichkeit auf organisationstheoretische Zusammenhänge rechtfertigungsbedürftig. Die Frage nach dem Zusammenhang unserer systematischen Forschungsbemühungen und den exegetisch relevanten Bezugspunkten wäre weiterführend genauer zu betrachten. Zum anderen basiert die systematische Perspektive auf einem Vorverständnis der Inkommensurabilität. Wir führten Inkommensurabilität als Sammelbegriff für ein Bündel von spezifischen Problemen ein. Dadurch verengten wir den Inkommensurabilitätsbegriff auf seinen problematischen Kern. Diese Verengung im Zuge der Rekonstruktion dieses Konzeptes ist laut Oberheim und Hoyningen-Huene (1997, S. 453) problematisch. Weiterführend wäre dem Umstand Rechnung zu tragen, dass Inkommensurabilität auch objekttheoretische Aspekte hat und dadurch Theorievergleiche im Feld der Organisationstheorie Eingang in die Überlegungen finden sollten.

Zweitens ist der Umgang mit der EOT als Transferobjekt

kritisch zu beleuchten. Wie mehrfach betont, handelt es sich bei der EOT um einen inhaltlich umfassenden und sehr ausdifferenzierten Ansatz. Der Umfang der Untersuchung machte eine Einschränkung des Textkorpus, der in die Reflexionen mit einbezogen werden konnte, unumgänglich. Im Rahmen einer weiterführenden Untersuchung wären insbesondere zwei Theorieteile der EOT vertieft zu analysieren. Zum einen handelt sich um die Ausführungen zur Frage nach der Möglichkeit von intersubjektiver Verständigung vor dem Hintergrund einer konstruktivistischen Perspektive. Die vorgetragene Argumentation hat deutlich gemacht, dass es sich beim Konzept der Intersubjektivität quasi um ein Gegenmittel für die Probleme handelt, welche an Inkommensurabilität gekoppelt sind. Die Relevanz einer derartigen Untersuchung ist somit direkt erkennbar. Zum anderen wäre die Auseinandersetzung der EOT mit der Theorie des kommunikativen Handelns von Habermas einzubeziehen (vgl. exemplarisch Kirsch, 1992, S. 31 ff.; Kirsch et al., 1999a, S. 41 ff., 92 ff., 120 ff.). Auch dieser Theoriebaustein ist im Hinblick auf die Frage nach der Möglichkeit von intersubjektiver Verständigung interessant.

Wenden wir uns abschliessend erneut unserem Befund zum Theoriepluralismus im Feld der Organisationstheorie zu. Wie vorab angedeutet, eröffnet unserer Ansicht nach die vorliegende Untersuchung den Raum für neue organisationstheoretische Forschungsbemühungen. Wie sind diese anzulegen? Die propagierte Forschungsstrategie möchten wir unter dem Label *„Mut zur philosophischen Reflexion!“* für die organisationstheoretische Grundlagendebatte vorschlagen. Philosophische Reflexion meint in diesem Zusammenhang etwas Konkretes. Sobald man sich innerhalb der Organisationstheorie auf der Ebene der Grundlagenfragen bewegt, ist es notwendig, die wissenschaftstheoretische Perspektive proaktiv einzunehmen und in der notwendigen Tiefe zu reflektieren. Nur wenn man bereit ist, sich genuin auf den State of the Art der philosophischen Debatte über die Möglichkeit und Grenzen menschlicher Erkenntnis, die Funktionsweise von Sprache und sozialer Interaktion einzulassen, kann ein konsistentes Fundament für die Organisationstheorie erarbeitet werden. Die vorliegende Untersuchung hat darüber hinaus deutlich gemacht, dass die wissenschaftstheoretischen Grundannahmen weitreichende Konsequenzen für Konstruktionsbemühungen in der Organisationstheorie haben.

Ein weiterführendes Anwendungsgebiet für diese Forschungsstrategie ist die Debatte um die Frage nach der Handhabung von Inkommensurabilität im Feld der Organisationstheorie. Die Vielzahl an vorgeschlagenen, konkreten Strategien (vgl. exemplarisch Lewis und Grimes, 1999; Scherer und Steinmann, 1999) ist erst dann kritisch zu würdigen, wenn man sich im Klaren darüber ist, basierend auf welchen wissenschaftstheoretischen Grundlagen man die Thematik der Inkommensurabilität betrachten möchte. In diesem Sinne: Mut zur philosophischen Reflexion!

Literatur

Astley, W. G. und van de Ven, A. H. Central Perspectives and Debates in Organization Theory. *Administrative Science Quarterly*, 28(2):245–273, 1983.

Backer, G. P. und Hacker, P. M. S. Wittgenstein - understanding & meaning. In Hacker, P. M. S., editor, *An analytical commentary on the "Philosophical Investigations"*, volume 1. Blackwell, Oxford, 2005.

Bird, A. *Thomas Kuhn*. Princeton University Press, Princeton, 2000.

Boyd, R. N. Reference, (in)commensurability and meanings. In Cohen, R., Renn, J., und Gavroglu, K., editors, *Incommensurability and related matters*, pages 1–63. Boston Studies in the Philosophy of Science, Volume 216. Dordrecht: Kluwer Academic Publishers, 2001.

Brown, H. I. Incommensurability and reality. In Cohen, R., Renn, J., und Gavroglu, K., editors, *Incommensurability and Related Matters*, pages 123–142. Boston Studies in the Philosophy of Science, Volume 216. Dordrecht: Kluwer Academic Publishers, 2001.

Burrell, G. Normal Science, Paradigms, Metaphors, Discourses and Genealogies of Analysis. In Clegg, S. R., Hardy, C., und Nord, W. R., editors, *Handbook of Organization Studies*, pages 642–658. Sage Publications, London, 1996.

Burrell, G. Organization Paradigms. In Sorge, A. und Malcom, W., editors, *The IEBM Handbook of Organizational Behaviour*, pages 33–47. International Thomson Business Press, London, 1997.

Burrell, G. und Morgan, G. *Sociological Paradigms and Organisational Analysis - Elements of the Sociology of Corporate Life*. Ashgate, Aldershot, 1979.

Carrier, M. Changing laws and shifting concepts. In Cohen, R., Renn, J., und Gavroglu, K., editors, *Incommensurability and related matters*, pages 65–90. Boston Studies in the Philosophy of Science, Volume 216. Dordrecht: Kluwer Academic Publishers, 2001.

Carrier, M. Grundproblem der Sprachphilosophie. Unveröffentlichtes Vorlesungsskript der Universität Bielefeld, 2008. URL http://www.uni-bielefeld.de/philosophie/personen/carrier/Sprachphilosophie.pdf.

Chen, X. Thomas Kuhn's latest notion of incommensurability. *Journal for General Philosophy of Science*, 28:257–273, 1997.

Donaldson, L. In Defence of Organization Theory - A reply to the critics. In Wedderburn, D., editor, *Management and industrial relation series*, volume 9. Cambridge University Press, Cambridge, 1985.

Evered, R. und Louis, M. Alternative Perspectives in the Organizational Sciences - "Inquiry from the Inside" and "Inquiry from the Outside". *Academy of Management Review*, 6(3):385–395, 1981.

Feyerabend, P. K. An Attempt at a Realistic Interpretation of Experience. *Proceedings of the Aristotelian Society New Series*, 58:143–170, 1958.

Feyerabend, P. K. Explanation, Reduction, and Empiricisms. In Feigl, H. und Maxwell, G., editors, *Scientific Explanation, Space, and Time*, pages 28–97. University of Minnesota Press, Minneapolis, 1962.

Feyerabend, P. K. Problems of Empiricism. In Colodny, P. K., editor, *Beyond the Edge of Certainty*, pages 145–260. Prentice-Hall, New Jersey, 1965.

Feyerabend, P. K. Explanation, Reduction and Empiricism. In Feyerabend, P. K., editor, *Realism, Rationalism and Scientific Method - Philosophical Papers*, volume 1, pages 44–96. Cambridge University Press, Cambridge, 1981.

Feyerabend, P. K. *Against Method*. Verso, London, 1993.

Glock, H. J. *A Wittgenstein Dictionary*. Blackwell Publishing, Oxford, 1996.

Glynn, M. A., Barr, P. S., und Dacin, M. T. Pluralism and the Problem of Variety. *The Academy of Management Review*, 25(4):726–734, 2000.

Hacking, I. *Einführung in die Philosophie der Naturwissenschaften*. Reclam, Stuttgart, 1996.

Hannan, M. T. und Freeman, J. *Organizational Ecology*. Harvard University Press, Cambridge, 1989.

Hanson, N. R. *Patterns of Discovery*. Cambridge University Press, Cambridge, 1965.

Hassard, J. Sociology and Organization Theory - Positivism, Paradigms and Postmodernity. In Brown, W., Hopwood, A., und Willman, P., editors, *Cambridge Studies in Management*. Cambridge University Press, Cambridge, 1993.

Hoyningen-Huene, P. *Die Wissenschaftsphilosophie Thomas S. Kuhns*. Vieweg, Braunschweig, 1989.

Hoyningen-Huene, P. Kuhn's Conception of Incommensurability-Kuhn's Linguistic Turn. *Studies in the History and Philosophy of Science*, 21(3):481–492, 1990.

Irzik, G. und Grünberg, T. Whorfian Variations on Kantian Themes - Kuhn's Linguistic Turn. *Studies in the History and Philosophy of Science*, 29(2): 207–221, 1998.

Jackson, N. und Carter, P. In Defence of Paradigm Incommensurability. *Organization Studies*, 12(1):109–127, 1991.

Kirsch, W. *Kommunikatives Handeln, Autopoiese, Rationalität - Sondierungen zu einer evolutionären Führungslehre*. Verlag Barbara Kirsch, München, 1992.

Kirsch, W. *Wegweiser zur Konstruktion einer evolutionären Theorie der strategischen Führung - Kapitel eines Theorieprojektes*. Verlag Barbara Kirsch, München, 1996.

Kirsch, W. Evolutionäre Organisationstheorie I - Fortsetzung eines Projektes der Moderne mit anderen (postmodernen?) Mitteln, 1998. URL http://www.zfog.bwl.uni-muenchen.de/veroffentlichungen/-eot/index.html.

Kirsch, W. und Guggemos, W. C. Evolutionäre Organisationstheorie II - Führung: ein erklärungsbedürftiges Phänomen. unveröffentlichter Arbeitstext, 1999b. URL http://www.zfog.bwl.uni-muenchen.de/veroffentlichungen/-eot/index.html. (besucht am 26.08.2010).

Kirsch, W. und Heeckt, N. N. Evolutionäre Organisationstheorie VI - Unternehmenspolitik, unveröffentlichter Arbeitstext. unveröffentlichter Arbeitstext, 2001. URL http://www.zfog.bwl.uni-muenchen.de/veroffentlichungen/-eot/index.html. (besucht am 26.08.2010).

Kirsch, W. und Weber, M. Evolutionäre Organisationstheorie IV - Themen und Agenden in Organisationen, 1999c. URL http://www.zfog.bwl.uni-muenchen.de/veroffentlichungen/-eot/index.html. (besucht am 26.08.2010).

Kirsch, W. und Weber, M. Evolutionäre Organisationstheorie V - Perspektiven einer Prozessorientierung, 1999d. URL http://www.zfog.bwl.uni-muenchen.de/veroffentlichungen/-eot/index.html. (besucht am 26.08.2010).

Kirsch, W., Brunner, K., und Eckert, N. Evolutionäre Organisationstheorie III - "Baustellen" eines Theorieprogramms. unveröffentlichter Arbeitstext, 1999a. URL http://www.zfog.bwl.uni-muenchen.de/veroffentlichungen/-eot/index.html. (besucht am 26.08.2010).

Kirsch, W., Seidl, D., und van Aaken, D. *Betriebswirtschaftliche Forschung - Wissenschaftstheoretische Grundlagen und Anwendungsorientierung*. Schäffer-Poeschl, Stuttgart, 2007.

Kirsch, W., Seidl, D., und van Aaken, D. *Unternehmensführung. Eine evolutionäre Perspektive*. Schäffer-Poeschl, Stuttgart, 2009.

Kuhn, T. S. *The Structure of Scientific Revolutions*. The University of Chicago Press, Chicago, 1962.

Kuhn, T. S. Objectivity, Value Judgement and Theory Choice. In Kuhn, T. S., editor, *The Essential Tension - Selected Studies in Scientific Tradition and Change*, pages 320–339. The University of Chicago Press, Chicago, 1977a.

Kuhn, T. S. Preface. In Kuhn, T. S., editor, *The Essential Tension - Selected Studies in Scientific Tradition and Change*, pages 9–23. The University of Chicago Press, Chicago, 1977b.

Kuhn, T. S. Second Thoughts on Paradigms. In Kuhn, T. S., editor, *The Essential Tension - Selected Studies in Scientific Tradition and Change*, pages 293–319. The University of Chicago Press, Chicago, 1977c.

Kuhn, T. S. The Essential Tension: Tradition and Innovation in Scientific Research. In Kuhn, T. S., editor, *The Essential Tension - Selected Studies in Scientific Tradition and Change*, pages 225–239. The University of Chicago Press, Chicago, 1977d.

Kuhn, T. S. Commensurability, Comparability, Communicability, Proceedings of the Biennial Meeting of the Philosophy of Science Association. 2:669–688, 1982.

Kuhn, T. S. *The Structure of Scientific Revolutions*. The University of Chicago Press, Chicago, 1996.

Lakatos, I. Falsification and the Methodology of Scientific Research Programmes. In Lakatos, I. und Musgraves, A., editors, *Criticism and the Growth of Knowledge*, pages 91–195. Cambridge University Press, Cambridge, 1970.

Lewis, M. W. und Grimes, A. J. Metatriangulation - Building Theory from Multiple Paradigms. *The Academy of Management Review*, 24(2):672–690, 1999.

Lorenz, K. Relation. In Mittelstrass, J., editor, *Enzyklopädie Philosophie und Wissenschaftstheorie*, pages 552–555. Verlag J. B. Metzler, Stuttgart, 2004.

Luhmann, N. *Soziale Systeme - Grundriss einer allgemeinen Theorie*. Suhrkamp, Frankfurt am Main, 1984.

Macdonald, G. Alfred Jules Ayer. Stanford Encyclopedia of Philosophy, Sum-

mer 2010 Edition, 2005. URL http://plato.stanford.-edu/entries/ayer/. (besucht am 26.08.2010).

McKinley, W. und Mone, M. A. The Re-Construction of Organization Studies - Wrestling with Incommensurability. *Organization*, 5(2):169–189, 1998.

McKinley, W., Mone, M. A., und Moon, G. Determinants and Development of Schools in Organization Theory. *Academy of Management Review*, 24 (4):634–648, 1999.

Morgan, G. Paradigms, Metaphors, and Puzzle Solving in Organization Theory. *Administrative Science Quarterly*, 25(4):605–622, 1980.

Morgan, G. Paradigm diversity in organizational research. In Hassard, J. und Pym, D., editors, *The theory and philosophy of organizations - Critical issues and new perspectives*, pages 13–29. Routledge, London, 1990.

Nagel, R. Strategieberatung - Expertenorientiert oder systemisch. *Hernsteiner*, 4:14–20, 2001.

Nicolai, A. T. und Seidl, D. That's relevant! Towards a taxonomy of practical relevance. *Organization Studies*, 31, im Druck.

Oberheim, E. und Hoyningen-Huene, P. Incommensurability, Realism and Meta-Incommensurability. *Theoria Segunda Epoca*, 12(3):447–465, 1997.

Oberheim, E. und Hoyningen-Huene, P. The Incommensurability of Scientific Theories. Stanford Encyclopedia of Philosophy, Summer 2010 Edition, 2009. URL http://plato.stanford.edu/entries/incommensurability/. (besucht am 26.08.2010).

Perovich, A. N. Inkommensurabilität - ihre Unterarten und ontologischen Konsequenzen. In Duerr, H. P., editor, *Versuchungen - Aufsätze zur Philosophie Paul Feyerabends*, volume 2, pages 76–94. Suhrkamp, Frankfurt am Main, 1981.

Pfeffer, J. Barriers to the Advance of Organizational Science - Paradigm Development as a Dependent Variable. *The Academy of Management Review*, 18(4):599–620, 1993.

Picot, A., Dietl, H., und Franck, E. *Organisation - Eine ökonomische Perspektive*. Schäffer-Poeschl, Stuttgart, 2008.

Preston, J. *Kuhn's The Structure of Scientific Revolutions - A Reader's Guide*. Continuum, New York, 2008.

Quine, W. V. O. *Grundzüge der Logik*. Suhrkamp, Frankfurt am Main, 1969.

Roethlisberger, F. J. und Dickson, W. J. *Management and the Worker - An Account of a Research Program Conducted by the Western Electric Company Howthorne Works Chicago*. Harvard University Press, Cambridge, 1964.

Sankey, H. und Hoyningen-Huene, P. Introduction: Incommensurability and related matters. *Boston Studies in the Philosophy of Science*, 216:7–34, 2001.

Sankey, H. *The Incommensurability Thesis*. Ashgate, Aldershot, 1994.

Scheffler, I. *Science and Subjectivity*. The Bobbs-Merrill Company, Inc., Indianapolis, 1967.

Scherer, A. G. und Dowling, M. J. Towards a Reconciliation of the Theory-Pluralism in Strategic Management - Incommensurability and the Constructivist Approach of the Erlangen School. In Shrivastava, P. und Stubbart, C., editors, *Advances in Strategic Management - Challenges from outside the Mainstream*, volume 12, pages 195–247. Jai Press Inc., London, 1995.

Scherer, A. G. und Steinmann, H. Some Remarks on the Problem of Incommensurability in Organization Studies. *Organization Studies*, 20(3): 519–544, 1999.

Schulte, J. *Wittgenstein - Eine Einführung*. Reclam, Stuttgart, 2001.

Simon, F. B. *Einführung in die systemische Organisationstheorie*. Carl-Auer-Systeme, Heidelberg, 2007.

Sorge, A. Organization Behaviour. In Sorge, A. und Malcom, W., editors, *The IEBM Handbook of Organizational Behaviour*, pages 3–20. International Thomson Business Press, London, 1997.

Tadajewski, M. The Debate That Won't Die? Values Incommensurability, Antagonism and Theory Choice, Organization. 16(4):467–185, 2009.

Thommen, J.-P. und Achleiter, A.-K. Allgemeine Betriebswirtschaftslehre: Umfassende Einführung aus managementorientierter Sicht. *Wiesbaden: Gabler*, 2006.

Uebel, T. Vienna Circle. Stanford Encyclopedia of Philosophy, Summer 2010 Edition, 2006. URL http://plato.stanford.edu/entries/-vienna-circle/. (besucht am 26.08.2010).

von Foerster, H. Über das Konstruieren von Wirklichkeiten. In von Foerster, H., editor, *Heinz von Foerster - Wissen und Gewissen, Versuch einer Brücke*, pages 25–49. Suhrkamp Verlag, Frankfurt am Main, 1993.

von Foerster, H. und Pörksen, B. *Wahrheit ist die Erfindung eines Lügners - Gespräche für Skeptiker*. Carl Auer Syteme, Heidelberg, 2003.

von Glaserfeld, E. Konstruktion der Wirklichkeit und des Begriffs der Objektivität. In von Siemens Stiftung, C. F., editor, *Einführung in den Konstruktivismus*, pages 1–26. R. Oldenbourg Verlag, München, 1985.

von Glaserfeld, E. *Radikaler Konstruktivismus - Ideen, Ergebnisse, Probleme*. Suhrkamp, Frankfurt am Main, 1998.

Weaver, G. R. und Gioia, D. A. Paradigms Lost - Incommensurability vs. Structurionist Inquiry. *Organization Studies*, 15(4):565–590, 1994.

Williamson, O. E. The Logic of Economic Organization. In Williamson, O. E. und Winter, S. G., editors, *The Nature of the Firm - Origins, Evolution, and Development*, pages 90–116. Oxford University Press, Oxford, 1993.

Winter, W. *Theorie des Beobachters - Skizzen zur Architektonik eines Metatheoriesystems*. Verlag Neue Wissenschaft, Frankfurt am Main, 1999.

Wittgenstein, L. Philosophische Untersuchungen. In Wright, G. H. und Anscombe, G. E. M., editors, *Ludwig Wittgenstein - Werkausgabe Band 1*, volume 1. Suhrkamp, Frankfurt am Main, 1999.

Junior Management Science 1 (2016) 1-300

Junior Management Science

journal homepage: www.jums.academy

Effektive Prävention ethisch-moralisch bedingter Unternehmenskrisen – Eine fallbasierte Untersuchung in Deutschland und den USA

Sonja Anna-Maria Hammel
Ludwig-Maximilians-Universität München

Abstract

Unternehmen geraten schon lange nicht mehr nur mit positiven Meldungen in die Schlagzeilen der Presse. Unternehmen sind Organisationen, die sich aus Menschen zusammensetzen. Menschen treffen zuweilen ethisch-moralisch bedenkliche Entscheidungen, lassen sich z.B. zu korruptem oder betrügerischem Verhalten hinreißen. Solch ein Verhalten kann sich, auf die gesamte Organisation bezogen, über Prozesse positiver Rückkopplungseffekte, zu starren Strukturen und festgefahrenen Verhaltensmustern entwickeln. Einmal festgefahrene Strukturen können auf Dauer nur noch durch ein Aufrütteln des ganzen Unternehmens, z.B. in Form einer Unternehmenskrise, aufgebrochen werden. Diese Schilderung spiegelt eine ethisch-moralisch bedingte Unternehmenskrise aus der Perspektive der Theorie Organisationaler Pfadabhängigkeit wider.

Mittels einer fallbasierten Untersuchung selektierter DAX 30- und Dow Jones 30-Krisenfälle im Zeitraum von 1988 bis 2014 schafft diese explorative Arbeit ein tieferes Verständnis ethisch-moralisch bedingter Unternehmenskrisen. Eine integrierte Betrachtung ethisch-moralisch bedingter Unternehmenskrisen und der Theorie Organisationaler Pfadabhängigkeit soll Ansatzpunkte für eine effektive Prävention dieser Krisen liefern. Zu diesem Zweck wird eine vergleichende Analyse besonders für diese Art von Unternehmenskrisen anfälliger bzw. unanfälliger Unternehmen durchgeführt. Dabei werden Maßnahmen wie Reputationsmanagement, Corporate Governance oder Compliance-Programme etc. auf ihre Eignung als Krisenpräventionsmaßnahmen untersucht.

Die vorliegende Arbeit liefert wertvolle Erkenntnisse zum Verständnis ethisch-moralisch bedingter Unternehmenskrisen. Abgeleitet werden darüber hinaus qualitative Ansätze für ein Pfadbrechen, d.h. das Nutzen einer solchen Unternehmenskrise als Chance, um mit festgefahrenen Strukturen und Verhaltensweisen, also dem Pfad, zu brechen, um künftige ethisch-moralisch bedingte Unternehmenskrisen zu vermeiden.

Keywords: Prävention von Unternehmenskrisen, Krisenmanagement, Organisationale Pfadabhängigkeit, Qualitative Untersuchung (USA/ Deutschland)

1. Einleitung und Ziele der wissenschaftlichen Arbeit

> *"When written in Chinese, the word crisis is composed of two characters – one represents danger, and the other represents opportunity."* (John F. Kennedy, 1959 (Kennedy, 2014))

Dieses Zitat stammt aus einer Ansprache des US-amerikanischen Präsidenten John F. Kennedy, die er 1959 am United Negro College Fund in Indianapolis hielt. Die Botschaft ist offensichtlich: Eine Krise birgt nicht nur Gefahren. Eine Krise kann eine Chance bzw. eine Möglichkeit sein, neue und bessere Pfade einzuschlagen.

Diese Botschaft würden auch einige der großen Unternehmen unterschreiben, die schon einmal eine Unternehmenskrise durchlaufen haben. In den ersten Phasen der Krise, im Zuge des ersten Aufwallens eventueller Entdeckungen, Anschuldigungen oder sonstiger prekärer Enthüllungen, dominiert sicherlich die Wahrnehmungskomponente einer Krise als Gefahr. Werden jedoch wertvolle Erkenntnisse im Hinblick auf künftige Präventionsmaßnahmen aus solchen Vorkommnissen gezogen, so verlagert sich die Wahrnehmung der Krise möglicherweise hin zur positiven Sicht: Die Krise als Chance.

Ob aktuell der ADAC oder aus vergangenen Jahren Siemens, MAN oder VW – es scheint als sei jedes große Unternehmen zumindest einmal Opfer seiner eigenen Vergangenheit geworden. Am Anfang dieser Unternehmenskrisen

DOI: http://dx.doi.org/10.5282/jums/v1i1pp216-274

standen häufig ethisch-moralisch bedenkliche Entscheidungen entsprechender Entscheidungsträger wie Ressortvorstände oder andere Verantwortliche. Andererseits existieren Unternehmen, um die es krisentechnisch sehr still ist, die ohne in der Öffentlichkeit diskutierte Krisen ihren Platz in der Unternehmenslandschaft behaupten. Die Frage nach den Unterschieden solcher Unternehmen, die eher häufig aufgrund ethisch-moralisch bedenklicher Handlungen in der Öffentlichkeit präsent sind und solchen, die entweder über sehr große Zeiträume krisenfrei sind oder als Reaktion auf eine Krise Maßnahmen ergriffen haben, sodass sie darauffolgend krisenfrei sind, indiziert ein spannendes und relativ wenig untersuchtes Forschungsgebiet.

An dieser Stelle soll die vorliegende Arbeit ansetzen. Es gilt herauszufinden, wie sich relativ krisenunanfällige Unternehmen in ihren Krisenpräventionsmechanismen von den eher häufig krisengeplagten Unternehmen unterscheiden. Des Weiteren wird der Frage potenzieller Präventionsmechanismen ethisch-moralisch bedingter Unternehmenskrisen nachgegangen, welche bereits andere Autoren für Unternehmenskrisen im Allgemeinen erörtert haben (vgl. z.B. Coombs und Holladay, 2002).

Untersucht werden ethisch-moralisch bedingte Unternehmenskrisen vor dem theoretischen Hintergrund der Pfadabhängigkeit. Der Schwerpunkt der Pfadabhängigkeitsforschung beschäftigt sich im Ursprung mit technologisch orientierten Märkten und Unternehmen. Es existiert jedoch auch einige auf die Sozialwissenschaften angewandte Forschungsliteratur (vgl. z.B. Schreyögg et al., 2003; Sydow et al., 2009), auf welche diese Arbeit aufbaut. Bezugnehmend auf die Theorie pfadabhängiger Unternehmen wird untersucht, ob die analysierten deutschen und amerikanischen Fälle ethisch-moralisch bedingter Unternehmenskrisen Kennzeichen einer Pfadabhängigkeit aufweisen. Diese Untersuchung soll beantworten, ob und inwieweit pfadabhängige Prozesse ethisch-moralisch bedingte Unternehmenskrisen beeinflussen. Einen Schritt weitergehend wird beleuchtet, inwiefern sich Mechanismen wie Reputationsmanagement oder Corporate Governance Strukturen eignen, um mit „eingebrannten Strukturen“ zu brechen und damit effektive Krisenprävention zu betreiben.

Nach der Einleitung und der Darlegung der Ziele der Arbeit in Abschnitt eins, widmet sich Abschnitt zwei der theoretischen Fundierung. Zum einen wird das Verständnis einer ethisch-moralisch bedingten Unternehmenskrise hergeleitet. Zum anderen soll ein fundiertes Verständnis Organisationaler Pfadabhängigkeit nach Sydow et al. (2009) geschaffen werden. Abschnitt zwei liefert abschließend eine Idee der Integration der bislang kaum gemeinsam untersuchten Forschungsgebiete zu Unternehmenskrisen einerseits und Organisationaler Pfadabhängigkeit andererseits.

Der dritte Abschnitt erzählt zu Beginn die methodische Entstehungsgeschichte dieser Arbeit. Er liefert eine strukturierte Beschreibung des Vorgehens bei der fallbasierten Analyse relevanter Krisenfälle. Es wird dargestellt, wie die relevanten Krisenfälle identifiziert und anschließend kategorisiert und bewertet wurden. In Abschnitt drei werden daran anschließend die Fälle ethisch-moralisch bedingter Unternehmenskrisen für Deutschland anhand der DAX 30- und für die USA anhand der Dow Jones 30-Unternehmen untersucht. Schwerpunkt der Analyse ist dabei herauszufinden, in welchen Maßnahmen sich krisenanfällige von relativ krisenresistenten Unternehmen unterscheiden. Ebenso wird in diesem Kapitel ein Bezug zum Fundament der Pfadabhängigkeitstheorie hergestellt. Die identifizierten Krisenfälle werden dahingehend untersucht, inwieweit sich bei den krisenintensiven Unternehmen Hinweise auf eine historisch gewachsene Pfadabhängigkeit finden lassen. Der Abschnitt endet mit einer vergleichenden Betrachtung deutscher und amerikanischer Ergebnisse.

In Abschnitt vier werden die eigenen Ergebnisse noch einmal zusammengefasst dargestellt, strategische Implikationen diskutiert, Anregungen für weitere Forschungsanstrengungen gegeben sowie Limitationen der vorliegenden Arbeit aufgezeigt. Kapitel fünf rundet die Arbeit mit einer Konklusion ab.

2. Theoretische Einbettung des Forschungsthemas

Die vorliegende Arbeit liefert einen Beitrag zur Integration bisher unabhängiger Forschungsfelder. Um die Integration sukzessive aufzubauen, sollen zuerst beide Themenbereiche einzeln vorgestellt werden. Abschnitt 2.3 widmet sich abschließend einer Integrationsidee beider Forschungsfelder.

2.1. Forschung zu Unternehmenskrisen

Das Feld der Unternehmenskrisen wurde bereits in der ersten Hälfte des 20. Jahrhunderts etabliert und erlebte zu Zeiten der Weltwirtschaftskrise einen ersten Höhepunkt. Eine umfangreiche Auseinandersetzung erfuhr das wissenschaftliche Forschungsfeld jedoch im Zeitraum der zweiten Hälfte der 1970er bis in die Mitte der 1980er Jahre (vgl. Rüsen, 2009: S. 44). Im Folgenden wird zuerst der allgemeine Forschungsstand, bzw. ein themenrelevanter Auszug, wiedergegeben, wobei nach einer Definition vor allem auf mögliche Kategorisierungsschemata von Arten und Ursachen einer Unternehmenskrise eingegangen wird. Anschließend wird der Fokus auf ethisch-moralisch bedingte Unternehmenskrisen erläutert.

2.1.1. Definition, Arten & Ursachen einer Unternehmenskrise

Definition

Die Vielfalt an Definitionen einer Unternehmenskrise spiegelt die oft unterschiedlichen Forschungsstränge und –Schwerpunkte zu diesem Thema wider (vgl. z.B. Shrivastava und Mitroff, 1987: S. 5, Mitroff et al., 1987, Sohn und Lariscy, 2014: S. 23). Definitionen wachsen häufig auch in Abhängigkeit von der behandelten Phase einer Unternehmenskrise innerhalb des „Krisenprozesses“. Je nach betrachteter Phase betonen die aus den jeweiligen wissenschaftlichen Werken stammmenden Definitionen andere Teilaspekte einer Krise.

Coombs (2007a) charakterisiert, eine Unternehmenskrise als Prozess darstellend, einen „Crisis Management Life Cycle“, welcher drei Phasen einer Unternehmenskrise unterscheidet (siehe Abbildung 1). Hinter dieser Strukturierung verbirgt sich das Verständnis, dass in den verschiedenen Phasen einer Unternehmenskrise unterschiedliche Maßnahmen angemessen sind. Eine Unternehmenskrise in dieser Arbeit soll verstanden werden als eine „major catastrophe that may occur [either naturally or] as a result of human error, intervention or even malicious intent. It can include tangible devastation, such as the destruction of lives or assets, or intangible devastation, such as the loss of an organization's credibility or other reputational damage” (Argenti, 2007: S. 213). Für Argenti weist eine Krise damit einen unvorhersehbaren Charakter auf. Eine Abschätzung der Entwicklung einer Unternehmenskrise wird als extrem schwierig erachtet. Neben Ursachen und der Entwicklung einer Unternehmenskrise werden auch mögliche Folgen in der Definition erwähnt. Diese können „greifbar“ oder „ungreifbar“ sein – sie sind demnach keineswegs immer messbar, denn auch Schäden der Unternehmensreputation sind potenzielle Folgen. Unternehmensreputation meint in diesem Zusammenhang die „aggregate evaluation stakeholders make about how well an organization is meeting stakeholder expectations based on its past behaviors” (Coombs, 2007b: S. 164).

Wie bereits der Titel der Arbeit zum Ausdruck bringt, liegt das Augenmerk, bezugnehmend auf Coombs Crisis Management Life Cycle, vor allem auf der ersten Phase einer Unternehmenskrise. Um jedoch ein ganzheitliches Bild des von Coombs charakterisierten Crisis Management Life Cycle und ein detaillierteres Verständnis der Pre-Crisis Phase zu bekommen, wird nun ein kurzer Überblick der einzelnen Phasen gegeben.

Pre-Crisis

Unternehmenskrisen weisen meist einen prozesshaften Charakter auf. Sie sind Entwicklungen, die sich durch erste Anzeichen ankündigen und mit der Zeit ihren vernichtenden Charakter entfalten. Je später eine Krise erkannt wird, umso größer ist die Gefahr, die durch sie droht. Positiv formuliert bedeutet dies, dass bei frühem Erkennen durch geeignete Maßnahmen gegengelenkt werden kann. Eventuell wird das Unternehmen durch eine Krise gestärkt hervorgehen (vgl. Jossé und Ziegelmeyer, 2013: S. 80). Diese Pre-Crisis Phase beschreibt alle Aspekte einer Krise bzw. einer Krisenkommunikation, welche der eigentlichen Krise vorausgehen. Aufgezeigte Handlungsmöglichkeiten beziehen sich dabei hauptsächlich auf drei Felder (vgl. Coombs, 2007a: S. 2):

Im Zuge der signal detection entwickelt die Organisation ein Bewusstsein für eine potenzielle Krise. Die Analyse relevanter Informationen ermöglicht es der Organisation, Anzeichen einer potenziellen Krisensituation zu erkennen. Mithilfe des Bewusstseins über Warnsignale können manche Krisensituationen frühzeitig erkannt und bestenfalls vermieden werden (vgl. Gammelmark, 2011: S. 9).

Hat das Unternehmen eine potenzielle Krise anhand bestimmter Warnsignale erkannt, sollten schnellstmöglich (auf Stufe 2) alle zur Prävention nötigen Maßnahmen ergriffen werden, um die potenziellen Gefahren der durch die Warnsignale identifizierten Situation zu reduzieren/ abzuwenden. Eine effektive Krisenvermeidung stellt den Idealfall dar. Es gelingt einem Unternehmen nicht immer, eine sich abzeichnende Unternehmenskrise zu vermeiden. Hier setzt die dritte Stufe an – die Krisenvorbereitung. Auf dieser Stufe geht es vor allem um das Bewusstsein des Unternehmens, dass nicht alle Krisen vermeidbar sind. Diese Einsicht macht es jedoch umso nötiger, stets auf eine Unternehmenskrise vorbereitet zu sein (vgl. Gammelmark, 2011: S. 9).

Crisis

Die Phase der eigentlichen Krise besteht aus den zwei Stufen:

Der erste Schritt eines Unternehmens in der Krise ist ihre Anerkennung. Neben allen Warnsignalen wird dem Unternehmen in dieser Situation klar, dass es sich in der Krise befindet. Zu diesem Zeitpunkt erkennen auch Stakeholder, dass das Unternehmen in der Krise steckt und es gilt nun Schadensbegrenzung zu betreiben, was zur zweiten Stufe führt. An dieser Stelle sollte das Unternehmen alle nötigen und möglichen Handlungen ergreifen, um die Krisensituation einzudämmen. Dabei ist es u.a. wichtig, die Art der Krise (dazu mehr unter Abschnitt 2.1.1.2) zu berücksichtigen, um entsprechend passende Gegenmaßnahmen zu ergreifen (vgl. Gammelmark, 2011: S. 10).

Post-Crisis

In dieser Phase kehrt das Unternehmen zum Alltagsgeschäft zurück, jedoch muss es sich zuerst von allen (Nach-) Wirkungen der Krise erholen (1. Stufe recovery), um in einer nächsten Stufe die wichtigsten „Lessons Learned“ aus der bewältigten Krise abzuleiten (vgl. Coombs, 2007a: S. 9). Ergebnisse dieses Nachbearbeitungsprozesses sind Erkenntnisse über mögliche Vorbereitungsmaßnahmen für weitere potenzielle Krisen. Diese Phase dient also vor allem der Sicherstellung, dass die Krise tatsächlich beendet ist und die Unternehmensreputation nicht weiter durch die vergangene Krisensituation beschädigt wird (vgl. Gammelmark, 2011: S. 10).

Bezugnehmend auf die unterschiedlichen Phasen einer Unternehmenskrise konzentriert sich die Literatur vor allem auf die Untersuchung von Krisensituationen selbst bzw. auf die Phase danach. Eine Argumentation aus dieser retrospektiv-orientierten Perspektive vermag jedoch über einen geringfügigen Wert bezüglich der Verbesserung des Managements zukünftiger Krisen (vgl. Topper und Lagadec, 2013: S. 4). Betrachtet man Unternehmenskrisen jedoch wie Kahn et al. als „shock waves sent through systems“ (Kahn et al., 2013: S. 378), so bekommt der Auslöser solcher Schocks eine relativ größere Bedeutung im Hinblick auf die Prävention künftiger Krisen.

Arten und Verlauf

Auch bei der Typisierung von Unternehmenskrisen bietet die Literatur unzählige Möglichkeiten. Unterscheidungen finden sich gegliedert nach

- der rechtlichen Bedeutung (betriebswirtschaftliche vs. rechtliche Krise)

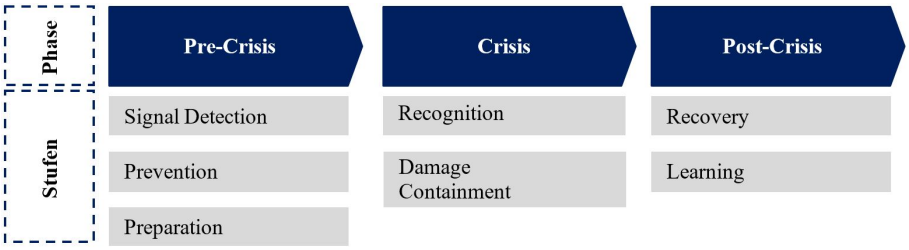

Abbildung 1: Crisis Management Life Cycle nach Coombs, 2007a (Quelle: eigene Darstellung)

- der Phase im Unternehmensentwicklungsprozess (Gründungskrise vs. Wachstumskrise vs. Stagnations-/ Schrumpfungskrise)
- dem Grad der Beeinflussbarkeit (potenzielle vs. latente vs. akute Krise)
- der Ursache (exogene vs. endogene Krise)
- dem auslösenden Unternehmensbereich (Krise einzelner leistungswirtschaftlicher Bereiche vs. finanzwirtschaftliche Krise vs. organisatorische Krise)
- den bedrohten Unternehmenspotentialen (Liquiditäts- vs. Erfolgs- / Rentabilitäts- vs. Produkt- / Absatz- vs. Strategische Krise) (vgl. LÜBECK, 2014 Steuerberater Rechtsanwälte GbR, 2014).

Im weiteren Verlauf wird sich vor allem auf die zuletzt genannte Kategorisierung bezogen, daher werden deren einzelne Stufen noch detaillierter erläutert. Eine Unternehmenskrise kann dabei mit einem Eisberg verbildlicht werden. Diese Assoziation bringt zum Ausdruck, dass dem Betrachter das eigentliche Ausmaß der Unternehmenskrise oft verborgen bleibt (vgl. Arnold, 2010: S.33), was in Abbildung 2 zu erkennen ist. **Sichtbare Krise**

Lange vor der akuten Krisengefährdung wird für Insider eine drohende Liquiditätskrise sichtbar. In Wirklichkeit ist die Liquiditätskrise das Ergebnis einer dahinter liegenden andauernden Erfolgskrise. Vor der Liquiditätskrise durchläuft ein Unternehmen eine Erfolgs-/ Rentabilitätskrise. Es kann die operativen Zahlungsverpflichtungen zwar noch erfüllen. Nachfragerückgänge, Preisverfall und Kostensteigerungen haben jedoch schon zu Gewinneinbrüchen oder Verlusten geführt, und notwendige Investitionen oder Sozialpläne können nicht mehr finanziert werden. Grund für eine Erfolgskrise kann eine Produkt- und Absatzkrise sein. Sie wird durch einen Rückgang der Umsatzerlöse, eine Zunahme der Vorratsbestände oder eine Unterauslastung der Produktionskapazitäten und einer damit einhergehenden Ergebnis- und Liquiditätsverschlechterung erkennbar (vgl. Arnold, 2010: S. 33).

Unsichtbare Krise

Eine Strategische Krise steht häufig am Beginn vieler Krisen und Insolvenzen. Typische Symptome sind eine anfängliche Fehlentscheidung, andauernde Konflikte, Polarisierungen und Reibungsverluste zwischen Gesellschaftern, Mitarbeitern etc. (vgl. Arnold, 2010: S. 33). So gibt es Schätzungen, die von einem Anteil der durch Strategische Krisen verursachten Insolvenzen bzw. Krisen von rund 60 Prozent ausgehen (vgl. Wöltje, 2011: S. 250). Die große Gefahr über eine Strategische Krise in die Insolvenz bzw. eine Krise zu geraten, besteht darin, dass häufig kein offensichtlicher Handlungsbedarf besteht (vgl. Wöltje, 2011: S. 252).

In Anlehnung an diese Typisierung von Unternehmenskrisen charakterisiert folgende Abbildung 3 den entsprechenden chronologischen Verlauf einer Krise.

Ursachen

Coombs bietet auch in Bezug auf mögliche Krisenursachen ein Kategorisierungsschema an, auf welches sich an dieser Stelle bezogen werden soll. Der Autor charakterisiert in Anlehnung an die Attributions-Theorie drei Arten bzw. Ursachen von Unternehmenskrisen (vgl. Coombs, 2007a: S. 8), welche in Abbildung 4 dargestellt sind.

1) Die Ursachen werden hierbei unterschieden nach der Verantwortlichkeit einer Krise. Sogenannte „Victim Crisis" bezeichnen dabei die Ursachen von Krisen, welche das geringste Maß an Verantwortlichkeit seitens des Unternehmens aufweisen. Ob Naturkatastrophen oder fälschlich verbreitete Gerüchte, welche das Unternehmen in eine Krise führen können – die Verantwortlichkeit ist hier schwerlich beim Unternehmen selbst zu suchen (vgl. Coombs, 2007a: S. 8).

2) Im Falle von „Accident Crisis" kann dem Unternehmen schon eher eine partielle Verantwortlichkeit angelastet werden wie bspw. in durch technisches Versagen eingeläuteten Krisenfällen (vgl. Coombs, 2007a: S. 8).

3) Die letzte Ursachenkategorie der „Preventable Crisis" wird dem Unternehmen als System die Krise angelastet. Hier finden sich Krisenfälle bedingt durch menschliches Versagen, bewusste Manipulation oder gar Gesetzesvergehen (durch das Management), welche Stakeholder in Gefahr bringen (vgl. Coombs, 2007a: S. 8).

Coombs liefert neben dieser Kategorisierung ein ergänzendes Schema, welches zu weiteren Einsichten über die Entstehung einer Unternehmenskrise beitragen kann. Zu unterscheiden sind dabei zwei Dimensionen der Krisenentstehung:

- Internal vs. External
- Intentional vs. Unintentional

Diese Betrachtungsweisen eröffnen eine neue Perspektive auf Krisenfälle, da vor allem im Rahmen der hier durchgeführten fallbasierten Untersuchung das Ziel verfolgt wird, die Hintergründe einer Krise zu verstehen – liegen die Ursachen im

Abbildung 2: Arten von Unternehmenskrisen (Quelle: Arnold, 2010: S. 33)

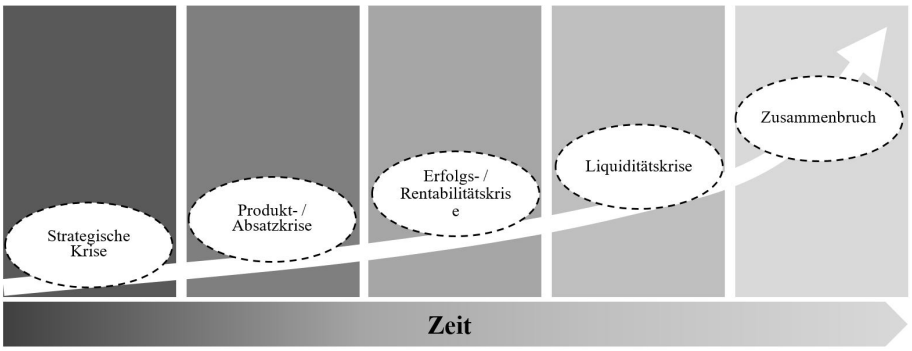

Abbildung 3: Verlauf einer Unternehmenskrise (Quelle: In Anlehnung an Wöltje, 2011: S. 251)

Unternehmen oder außerhalb? – und wurden krisenverantwortliche Handlungen absichtlich oder unabsichtlich vorgenommen? Zur Beantwortung dieser Fragen gibt Tabelle 1 eine Hilfestellung.

1) Im Falle eines Fauxpas ist umstritten, ob überhaupt von einer Krise gesprochen werden kann. Einige externe Gruppen/Personenkreise empfinden ein (möglicherweise ethisch-moralisches) Fehlverhalten des Unternehmens, jedoch kann keine Absichtlichkeit unterstellt werden (vgl. Coombs und Holladay, 1996: S. 285).

2) In Unternehmenskrisen, welche durch terroristische Anschläge bedingt sind, liegt die Krisenursache außerhalb des Unternehmens, wobei allerdings von einer absichtlichen Schädigung des Unternehmens durch externe Gruppen/Personenkreise auszugehen ist (vgl. Coombs und Holladay, 1996: S. 285).

3) Bei Unfällen jeglicher Art (menschliches oder technisches Versagen) liegt die Krisenursache zwar meist innerhalb des Unternehmens, jedoch kann hier von einer Nichtabsicht ausgegangen werden (vgl. Coombs und Holladay, 1996: S. 285).

4) In Verstoß-Fällen (Transgression) liegt die Krisenursache innerhalb des Unternehmens und zudem kann ein absichtliches (ethisch-moralisches) Fehlverhalten bzw. eine intentionale Schädigung des Unternehmens unterstellt werden (vgl. Coombs und Holladay, 1996: S. 285).

Aus der Vielfalt möglicher Typisierungen von Unternehmenskrisen wurden diese Schemata gewählt, da das Forschungsinteresse möglichen Präventionsmechanismen gilt und Coombs bereits in der Kategorisierung Anhaltspunkte für hier zu untersuchende Fälle (Preventable Crises: Strong Crisis Responsibility) gibt. Im Hinblick auf die hier durchgeführte fallbasierte Untersuchung ermöglicht die Vierfeldermatrix zudem eine weitere Fokussierung auf bestimmte Fälle in Abhängigkeit von Entstehungshintergründen der Krise.

2.1.2. Hier: Konzentration auf ethisch-moralisch bedingte Unternehmenskrisen

Bevor auf die konkrete Definition einer hier behandelten ethisch-moralisch bedingten Unternehmenskrise eingegangen wird, soll das zugrunde liegende Verständnis von Moral und Ethik kurz umrissen werden.

Im heutigen sprachlichen Gebrauch wird zwischen Moral und Ethik unterschieden. Die sprachlichen Wurzeln des Begriffs der Moral erforschend zeigt sich, dass das lateinische Wort mores die Übersetzung des griechischen Wortes ethos (Brauch, Gewohnheit, Charakter) ist. Gemeint ist damit

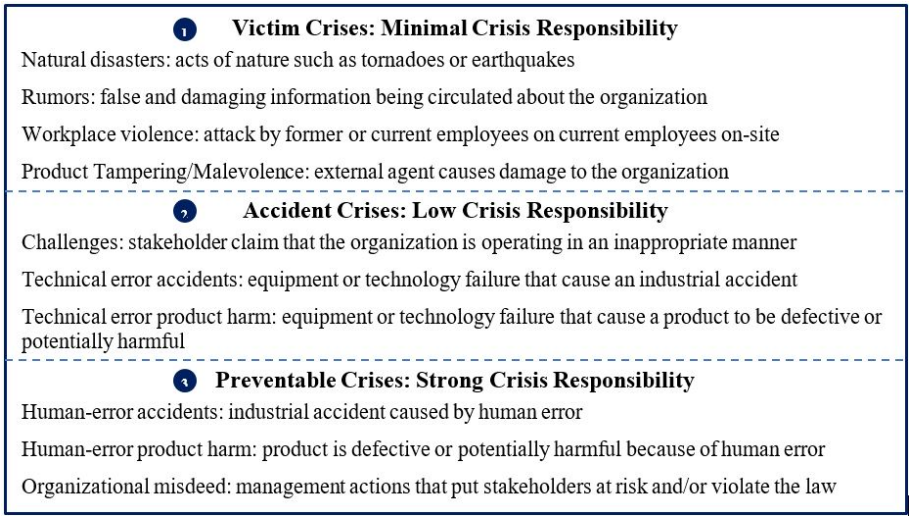

Abbildung 4: Crisis Types by Attribution of Crisis Responsibility (Quelle: Coombs, 2007a, S. 8, eigene Darstellung)

Tabelle 1: Characteristics of the crisis situation

(in Anlehnung an Coombs und Holladay, 1996, S. 284).

UNINTENTIONAL	INTENTIONAL	
1) Fauxpas	2) Terrorism	**EXTERNAL**
3) Accidents	4) Transgression	**INTERNAL**

die Einheitlichkeit der Verhaltensweisen von Gruppe auf Basis etablierter, seit Generationen gelebter Gewohnheiten. Moral ist demzufolge die „Gesamtheit aller moralischen Urteile, Normen, Ideale, Tugenden, Institutionen“ (Ricken, 1983: S. 18). Die Aufgabe der Ethik hingegen besteht in der Reflektion der Moral; sie ist die „philosophische Disziplin, die nach der Begründung der Moral fragt“ (Ricken, 1983: S. 18) und gilt damit als „Theorie des Richtigen und Guten“ (Mohr, 2010: S. 231). Beide Aspekte sollen in die Untersuchung hier analysierter Unternehmenskrisen einbezogen werden. Den Ursachen einer Krise können sowohl auf der Ebene tatsächlich beobachtbarer Handlungsweisen als auch in deren Verständnis und Begründung liegen.

Wie bereits im vorhergehenden Abschnitt dargelegt, zeigt die Literatur zum Kontext organisationaler Krisen eine enorme Vielzahl an Definitionen einer Krise. Die verschiedenen Typologien versuchen die Arten von Unternehmenskrisen zu kategorisieren bzw. zwischen ihnen zu differenzieren (vgl. Coombs und Holladay, 2002: S. 168). Die folgende Abbildung 5 greift auf die im vorherigen Abschnitt vorgestellten Kategorisierungen von Arten und Ursachen einer Unternehmenskrise zurück und zeigt die daraus abgeleitete Ausprägung einer Unternehmenskrise, auf welcher hier der Fokus liegt. Gesuchte und behandelte ethisch-moralisch bedingte Unternehmenskrisenfälle im Rahmen dieser explorativen Arbeit sind

- vom Grundsatz her Strategische Krisen, da sie mit einer wie auch immer gearteten Fehlentscheidung getriggered werden (vgl. Wöltje, 2011: S. 250),
- Preventable Crises (Strong Crisis Responsibility) der Art organizational misdeed nach Coombs (2007b), welche
- auf interne sowie absichtliche Entscheidungen von Organisationsmitgliedern zurückgehen, also Transgression-Fälle (vgl. Coombs und Holladay, 1996: S. 284). D.h. Unternehmenskrisen, die verursacht werden, da das „management knowingly violates laws or regulations“ (Coombs und Holladay, 2002: S. 171). In Abgrenzung zu den oben genannten Naturkatastrophen bezieht sich diese Definition auf Krisen mit internen Ursachen (vgl. Dubrovski, 2009: S. 39), welche sich im Kern auf jegliches bewusstes moralisches Fehlverhalten im Sinne eines Nichteinhaltens gesellschaftlicher, rechtlicher oder firmenspezifischer Richtlinien als Krisenursache bezieht.
- Behandelt werden management misconduct-Unternehmenskrisen im weiteren Sinne, losgelöst von der Ursachenbindung an Personen des Managements eines Unternehmens. Vielmehr soll es um ethisch-moralisch bedingte Unternehmenskrisen gehen, welche durch die Organisation als Ganze und/ oder das Management verursacht bzw. nicht verhindert werden. Wird eine Handlung unterlassen, so entspricht das

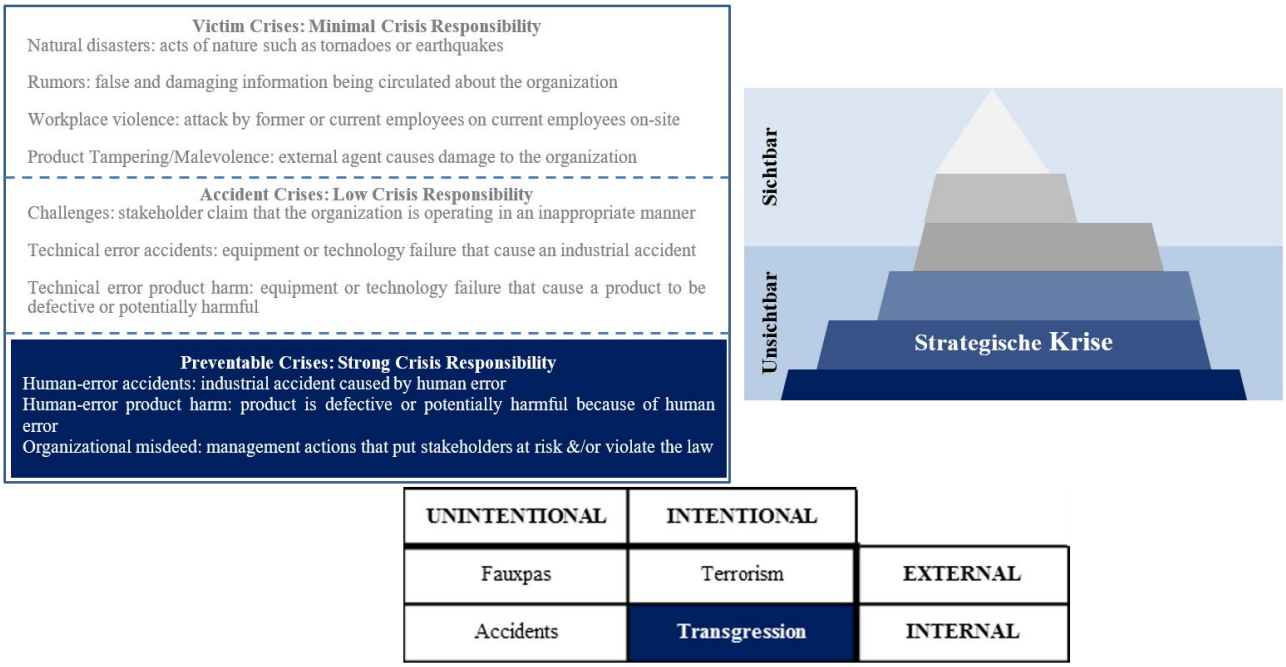

Abbildung 5: Visualisierung der behandelten Krisenform (Quelle: eigene Darstellung)

auch einer Entscheidung (vgl. Ricken, 1983: S. 13). Als Beispiele sind hier Krisen zu nennen, welche bspw. entstanden sind, weil Unternehmen Diskriminierung jeglicher Art oder menschenunwürdige Arbeitsbedingungen aktiv fördern oder passiv dulden. Ein solches Versagen und eine möglicherweise daraus resultierende ethisch-moralisch bedingte Unternehmenskrise sind nicht dem bewussten Fehlverhalten eines einzelnen Individuums anzulasten, sondern vielmehr als Versagen des gesamten Systems einzuordnen.

2.2. Forschung zur Pfadabhängigkeit von Unternehmen

Henning et al. (2013) untersuchen die empirische Literatur über Pfadabhängigkeit und Pfaderzeugung. Ihre Ergebnisse verdeutlichen, dass bzgl. des Verständnisses des Konzepts der Pfadabhängigkeit bis dato wenig Einheitlichkeit und Vergleichbarkeit bestehen (vgl. z.B. Dobusch und Kapeller, 2013: S. 288). Ihre Überprüfung zeigt, dass die Beliebtheit des Konzepts der Pfadabhängigkeit „nicht zu einer zielgerichteten Mehrung des Wissens über das Konzept, sondern vielmehr zu einer Kakofonie von Studien geführt hat“ (Henning et al., 2013: S. 1349). Diese Erkenntnis sollte bei der Auswertung und Deutung aller Studienergebnisse beachtet werden. Es lässt sich ein enormer Zuwachs der Publikationen zu diesem Forschungsstrang in den vergangenen Jahren feststellen. Abbildung 6 zeigt die Anzahl veröffentlichter Artikel (Web of Science Suchmaschine) bei der Eingabe der Suchwörter „path dependence“ oder „path dependency“. Die absolute Anzahl jährlich erschienener Artikel ist von neun im Jahr 1991 auf 57 in 2001 und 177 in 2011 gestiegen. Die hier dargestellte Anzahl an Artikeln bezieht sich auf alle Arten wissenschaftlicher Arbeiten – technologische, organisationale und industrielle Forschung eingeschlossen. Die Theorie der Pfadabhängigkeit erfreut sich wachsenden wissenschaftlichen Interesses. Jedoch konzentrieren sich die meisten Studien auf den Bereich der Technologien und des Innovation Managements (vgl. z.B. Ansari und Garud, 2009, Simmie, 2012a, Simmie, 2012b), was auf den Ursprung der Theorie zurückzuführen ist.

2.2.1. Definition und traditionelle Forschungsschwerpunkte

Die Literatur führt den theoretischen Kern des Begriffs auf den Ökonomen und Wirtschaftsmathematiker Arthur (1990, 1989) zurück. Bei Arthur wie bei anderen Autoren wird der Begriff der Pfadabhängigkeit herangezogen, um eine Dominanz bestimmter Produkte/ Standards über andere zu beschreiben (vgl. Henning et al., 2013: S. 1350). David (1985) lieferte bereits 1985 eine praktische Ausführung: die Etablierung der QWERTZ-Tastatur der Schreibmaschine. In den 1970/ 80er Jahren wurde der QWERTZ-Standard entwickelt und ist bis dato in seinen Grundzügen erhalten (vgl. Vergne, 2013: S. 1191). David stellte anhand dieser exemplarischen Analyse dar, dass andere Tastaturordnungen ebenso effizient oder gar effizienter sind, jedoch der „zufällige“ pfadabhängige Prozess dem Standard bis heute die Marktdominanz gewährleistet.

Andere Autoren bestreiten, dass es sich um eine zufällige Entwicklung handelte, denn „the most strategically savy entrepreneur, who also happens to be selling the best product, ends up winning the market competition“ (Vergne, 2013: S. 1191). Die Tastaturanordnung dominiert also nicht aufgrund zufälliger Entwicklungen, sondern wegen ihrer immer noch vorhandenen technischen und ergonomischen

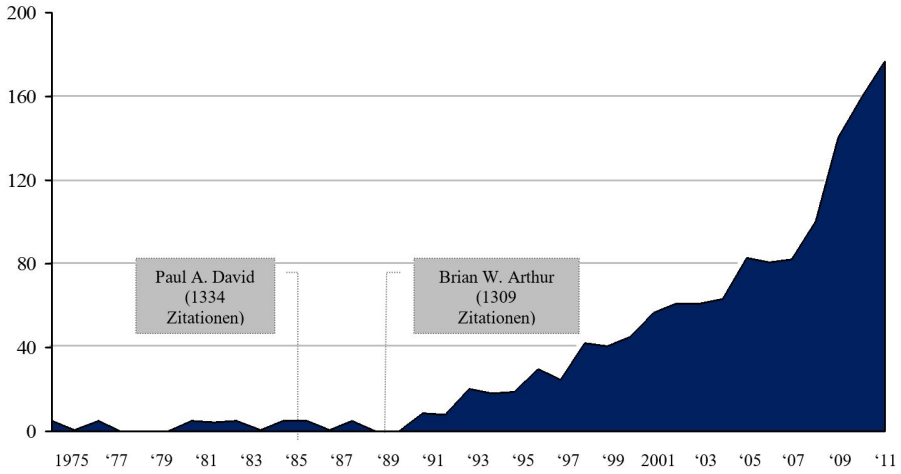

Abbildung 6: Anzahl veröffentlichter wissenschaftlicher Artikel zum Thema Pfadabhängigkeit, 1975–2011 (Quelle: Henning et al., 2013: S. 1350)

Überlegenheit (vgl. Kay, 2013: S. 1175).

Sind es also ausschließlich „historische Zufälligkeiten", welche den Pfad und damit die weitere Entwicklung eines Produkts/ einer Technologie gemäß dem Satz „History matters" determinieren? Dies ist eher zu verneinen, denn Pfadabhängigkeit ist weitaus mehr ist als eine reine Vergangenheitsabhängigkeit (vgl. z.B. Antonelli, 1997: S. 643f, Beyer, 2005: S. 5). Nur diese Einstellung zur Pfadabhängigkeit lässt eine Untersuchung wie eben diese hier zu. Denn wäre alles, was einem Unternehmen widerfährt, allein durch historische Entscheidungen determiniert, so hätte es als logische Konsequenz keinerlei Chance aus diesem Kreislauf auszubrechen, den Pfad zu verlassen oder eben eine Krise zu vermeiden (vgl. Antonelli, 1997: S. 643f). Darauf soll im weiteren Verlauf der Arbeit noch näher eingegangen werden.

Rückbeziehend auf die zwei Standardmodelle nach Arthur und David ist festzustellen, dass die Ansätze vor allem drei Gemeinsamkeiten aufweisen. Hennig et al. leiten daraus, wie im Folgenden zitiert, eine Art formale Interpretation der Pfadabhängigkeit ab (vgl. Henning et al., 2013: S. 1350f):

> 1. Improbable events with long-run effects on economic structures
>
> 2. Mechanisms such as increasing returns or network externalities persistently reinforce the situations created by chance events (often referred to as lock-in)
>
> 3. External shocks can disrupt [...] the locked-in pattern. Path dependence can be viewed as a non-ergodic process, or [...] a process where the current situation is dependent on its own history, but not to the extent that it eventually will, or can, return to its previous state.

Wie bereits angeführt, bekennt sich auch diese Interpretation zu den Wurzeln der Pfadabhängigkeit, welche in der produkt- bzw. technologieorientierten Forschung liegen. Eine erstmalige Anwendung in „sozialwirtschaftlich renovierter Form" (Sydow, 2010: S. 16) lieferten Thelen (1999), Pierson (2000), Mahoney (2000) auf politische Systeme und Ortmann (1995) erstmalig auf Organisationen. Die Theorie Organisationaler Pfadabhängigkeit ist heute integraler Bestandteil der Forschung rund um Unternehmen (vgl. Schreyögg et al., 2011: S. 81).

2.2.2. Theoretisches Rahmenwerk der Pfadabhängigkeit nach Sydow et al.

Sydow et al. (2009) bieten mit dem theoretischen Ansatz der Pfadabhängigkeit eine Grundlage, die auf dem Weg zu genau diesem Erkenntnisgewinn einen wertvollen Beitrag liefert. Der Ansatz beschäftigt sich im Kern nicht mit Unternehmenskrisen oder Präventionsmöglichkeiten. Eine Verbindung des theoretischen Ansatzes nach Sydow et al. und der Prävention von Unternehmenskrisen soll jedoch in dieser Arbeit hergestellt werden. Zentrales Konstrukt bei Sydow et al. ist die zuvor als allgemeines Konzept eingeführte Pfadabhängigkeit. Anhand ihres theoretischen Rahmenwerks analysieren die Autoren Ursachen und Entwicklungen, durch welche Unternehmen über verschiedene Phasen hinweg pfadabhängig werden. Was verstehen die Autoren jedoch unter Pfadabhängigkeit? Definiert wird Organisationale Pfadabhängigkeit als „a rigidified, potentially inefficient action pattern built up by the unintended consequences of former decisions and positive feedback processes" (Sydow et al., 2009: S. 696). Ein pfadabhängiges Unternehmen ist folglich gekennzeichnet durch (vgl. z.B. Dobusch und Schüßler, 2013, Schreyögg et al., 2003):

- eine gewisse Rigidität, Inflexibilität, Stabilitätsneigung,
- kollektiv ineffiziente oder suboptimal Zustände,
- festgefahrene (Verhaltens/Handlungs-) Muster, welche die

- unbeabsichtigten Konsequenzen vergangenheitsdeterminierter Entscheidungen sowie
- positiven Feedbacks sind.

Die Theorie Organisationaler Pfadabhängigkeit beruht auf einer internen Betrachtungsperspektive. Gesellschaftliche Parallelentwicklungen werden zunächst nicht berücksichtigt. Die folgende Abbildung 7 zeigt den mitunter langjährigen Prozess der Konstitution des organisationalen Pfads eines Unternehmens, welcher analytisch in drei Phasen untergliedert wird (vgl. Sydow, 2010: S. 24).

1) Zu Beginn jedes Pfadabhängigkeitsprozesses stehen eine oder mehrere Entscheidungen z.B. über diverse Handlungsalternativen. Die Autoren nennen diese kleinen, kontingenten, manchmal auch zufälligen Ereignisse „Small Events" (vgl. Sydow, 2010: S. 17). Phase I, die Preformations- oder auch Vorphase genannt, bezeichnet einen Status, in dem der Handlungsspielraum uneingeschränkt ist und jegliche Handlungsalternativen innerhalb der gesamten Variationsspanne gewählt werden können, wobei natürlich auch in dieser Situation die Handlungsmöglichkeiten von der Vergangenheit mitbestimmt sind. Eine einmal, aus welchen Beweggründen auch immer, gewählte Handlung mag sich dabei im Nachhinein als das dezisive Ereignis entpuppen, von dem ein sich selbst verstärkender Prozess seinen Ausgang nimmt. Die Autoren weisen darauf hin, dass in der Realität eine Vielzahl solcher kleiner Ereignisse zusammenwirken und als Pfad initiierende Ereignisse angesehen werden müssen (vgl. Sydow, 2010: S. 24). Am Ende dieses Zustands steht die sogenannte „critical juncture", ein Wendepunkt. Diese meint einen bestimmten Moment bzw. ein Ereignis, welches das Auftreten selbstverstärkender Effekte einläutet, die im weiteren Verlauf für die Persistenz verantwortlich sind (vgl. Sydow, 2010: S. 24 f.).

2) Die Formationsphase ist vor allem durch den stark verengten Handlungsspielraum gekennzeichnet, d.h. der eigentliche Pfad entwickelt sich hier. Sie beginnt mit der gerade vorgestellten critical juncture, von der ausgehend der sich selbst durch positive Rückkopplungen verstärkende Prozess beginnt. Wie entstehen diese selbstverstärkenden Effekte? Die Autoren betonen neben den ‚increasing returns' besonders vier Mechanismen. Sie können bspw. durch Koordinations- und Komplementaritätseffekte oder Lern- und Erwartungseffekte entstehen (vgl. Sydow et al., 2009: S. 699 ff.). Koch und Rothmann erweitern die Liste noch um Skalenerträge sowie direkte und indirekte Netzwerkexternalitäten (vgl. Koch und Rothmann, 2010: S. 82), wobei sich hier in der Literatur unzählige Fassungen finden. Als Resultat einer oder mehrerer dieser Effekte verfestigt sich bspw. eine Routine innerhalb der Organisation (-seinheit). Selbstverständlich sind einmal eingeschlagene Wege nicht alternativlos, jedoch bedarf es erheblicher und mit fortgeschrittener Pfadabhängigkeit größerer Anstrengung, einmal beschrittene Pfade zu verlassen (vgl. Sydow, 2010: S. 25). Dieser Prozessschritt der Pfadausbildung mündet in einen weiteren kritischen Punkt, den Lock-In. Dieser Punkt ist aus Prozesssicht sehr ausschlaggebend, da sich nun die dominante Alternative bzw. Handlungsweise durchsetzt und der folgende Verlauf determiniert ist (vgl. Sydow, 2010: S. 25 f.).

3) Die letzte Phase der Pfadabhängigkeit (Lock-In Phase) stellt eine Situation dar, in der kein Handlungsspielraum mehr besteht, allein die dominante Alternative wird ausgeführt (vgl. Schreyögg et al., 2003, Sydow et al., 2009: S. 694). Resultat: Das Unternehmen ist in seinem durch historische Entscheidungen eingeschlagenen Pfad „gefangen". Mag der Outcome des Unternehmens zu Beginn der Phase noch durch ein positives Lock-In getragen werden. Das Umschlagen des Lock-Ins in negative Wirkungen ist absehbar. Der dann erzielte Outcome ist u.U. ineffizient, da effizientere Alternativen aufgrund des Ausführens der dominanten Alternative bzw. Handlungsweise gar nicht in Betracht gezogen werden.

Die Kennzeichnen und Folgen jeder Phase finden sich in folgender Abbildung 8 zusammengefasst.

Laut den Autoren bilden selbstverstärkende Mechanismen das Herzstück Organisationaler Pfadabhängigkeit. Im ursprünglichen erforschten technologischen Umfeld der Theorie wurden diverse Mechanismen identifiziert, welche die Entwicklung und Ausbildung eines Pfads fördern. Diese Marktmechanismen sind nicht einfach auf Unternehmen übertragbar (vgl. Sydow et al., 2009: S. 698). Sydow et al. charakterisieren vor allem vier Mechanismen Organisationaler Pfadabhängigkeit:

Lerneffekte

Diese Effekte beruhen auf dem Verständnis, dass häufig wiederkehrende und auszuführende Aufgaben mit der Zeit verbessert und damit die Effizienz gesteigert wird. Dies hat abnehmende durchschnittliche Stückkosten pro Outputeinheit zur Folge. Neu zu erlernende Aufgaben werden als zunehmend aufwendig erachtet. Dies bedeutet, dass „the more attractive the chosen solution becomes because of accumulated skills and decreasing cost, the less attractive it is to switch to new learning sites" (Sydow et al., 2009: S. 700). Das Unternehmen verliert immer mehr die Fähigkeit nach frischen und neuen Alternativen Ausschau zu halten und damit kritisch mit etablierten Routinen umzugehen. Selbst neue Organisationsmitglieder scheitern mit eventuellen Verbesserungen und passen sich dem System bzw. dem Kodex an (vgl. Sydow et al., 2009: S. 700).

Komplementäre Effekte

Diese Effekte lehnen sich an die Funktionsweise der Economies of Scope an. Komplementaritäten bezeichnen Synergien, welche durch die Interaktion zweier oder mehrerer separater aber interkorrelierter Ressourcen, Regeln oder Verhaltensweisen erreicht werden. Der Vorteil wiederholter interkorrelierter Aktivitäten addiert sich nicht nur auf, sondern es entsteht ein zusätzlicher Mehrwert (vgl. Sydow et al., 2009: S. 699).

Koordinationseffekte

Diese Effekte beziehen sich auf die Vorteile regelgeleiteten Verhaltens von Organisationsmitgliedern und weist eine hohe Ähnlichkeit zum Prinzip der Economies of Scale auf. Denn „the more actors adopt and apply a specific institution, the more efficient the interaction among these actors is

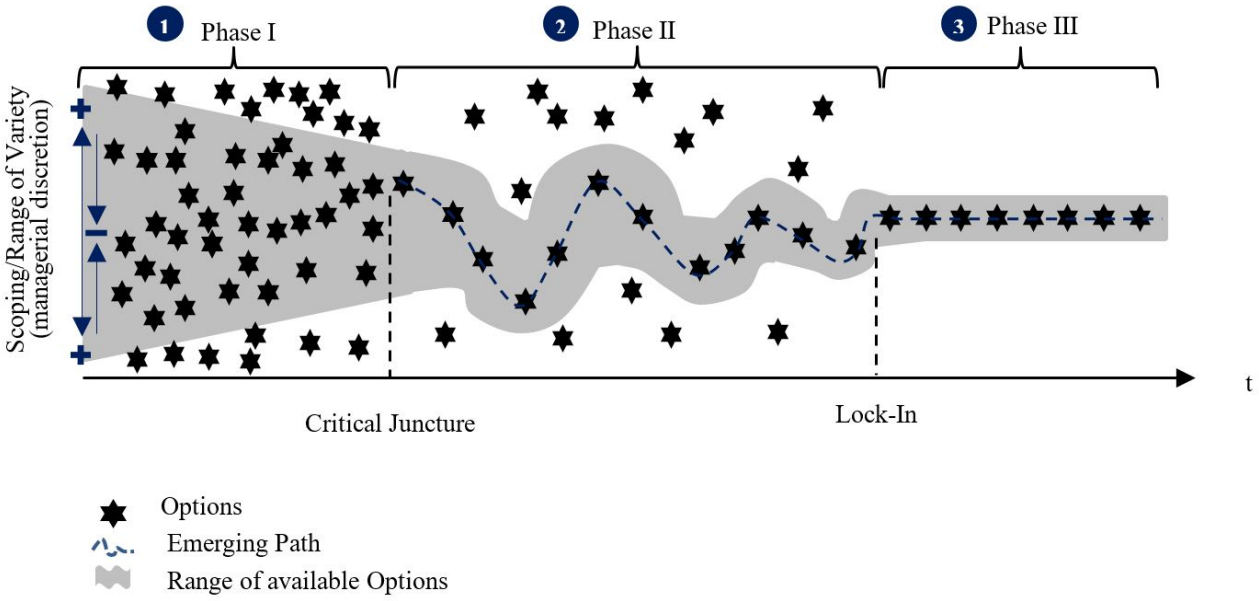

Abbildung 7: Die Konstitution und Entwicklung eines organisationalen Pfads (Quelle: Sydow et al., 2009: S. 692)

[...]"(Sydow et al., 2009: S. 699). Koordinationskosten können dadurch signifikant reduziert werden und für jedes einzelne Organisationsmitglied wird regelkonformes Verhalten zunehmend attraktiver. Durch die Vorteilhaftigkeit konformen Verhaltens resultiert dieser Prozess in einem gewissen festgefahrenen Verhalten (vgl. Sydow et al., 2009: S. 699).

Adaptive Erwartungseffekte

Adaptive Erwartungseffekte basieren auf der Annahme, dass Präferenzen nicht als fix angesehen werden, sondern Ergebnis eines interaktiven Prozesses in Abhängigkeit der Erwartungen anderer sind. Menschen streben danach, innerhalb des sozialen Gefüges auf der Gewinnerseite zu stehen und empfinden es als Belohnung, wenn andere dieselben Präferenzen teilen. Im organisationalen Kontext zeigt sich dieses Bestreben dadurch, dass Organisationsmitglieder bestimmte Routinen und Verhaltensweisen adoptieren und kontinuierlich ausführen in der Erwartung, dass „die anderen" dasselbe tun. Der Druck zu konformem Verhalten wächst durch die Angst des Outsider-Werdens (vgl. Sydow et al., 2009: S. 700).

2.3. Vereinbarkeit der Forschungsfelder zu Unternehmenskrisen und zur Pfadabhängigkeit

Jedes einzelne Forschungsfeld, das der Unternehmenskrisen und das der Pfadabhängigkeit, ist groß und birgt eine Vielzahl an Betrachtungsperspektiven. Und obwohl die beiden bei näherer Überlegung einige Ansatzpunkte für eine integrierte Betrachtung bieten, findet sich keine bzw. kaum wissenschaftliche Literatur, die sich um eine gemeinsame Betrachtung bemüht. Einzelne Arbeiten, welche eine integrative Untersuchung der Pfadabhängigkeit beim Auftreten von Unternehmenskrisen vornehmen, sind in Abbildung 9 zu sehen.

Tobias (2008) untersucht anhand eines Fallbeispiels den Beitrag der Theorie Organisationaler Pfadabhängigkeit für die Sanierungs- und Beratungspraxis. Er stellt dabei heraus, dass Unternehmen v.a. in Krisenzeiten an bewährten Handlungsmustern festhalten, wodurch sie eine sich abzeichnende Krise verschärfen. Der Autor zeigt auf, dass die Sanierung eines krisenbedrohten Unternehmens nur dann gelingen kann, wenn die Krisenpfade durchbrochen werden. Die Unternehmenskrise ist somit Anlass, mit persistenten und eventuell ineffizienten Verhaltensweisen zu brechen und so den Grundstein für den künftigen Unternehmenserfolg zu legen (vgl. Tobias, 2008: S. 119 ff.).

Lüttel (2009) untersucht in seiner Case Study einen historischen Auszug der Bankgesellschaft Berlin. Ihre Krise im Jahr 2001 führte sie nicht nur an den Rand einer Insolvenz, sondern löste auch einen politischen Skandal aus. Die Ursachen der Krise werden dabei zumeist auf die Inkompetenz einzelner Akteure oder gar krimineller Handlungen zurückgeführt. Lüttel erläutert die Entwicklungen vor dem konzeptionellen Hintergrund Organisationaler Pfadabhängigkeit. Es wird gezeigt, dass sich im Zeitverlauf eine sich selbstverstärkende Dynamik im strategischen Entscheidungsprozess entwickelte. Diese führte nicht nur zu einer sukzessiven Akkumulation von Verlustrisiken durch risikoreiche Geschäfte, sondern auch zu einem immer kleiner werdenden Entscheidungsspielraum. Erörtert wird also im Kern ein pfadabhängiger Prozess, der (zwangsläufig?) in einer Unternehmenskrise mündete (vgl. Lüttel, 2009: S. 17 ff.). Als Fallanalyse aufgebaut, verdeutlicht der Autor auf Grundlage zahlreicher journalistischer Recherchen, dass die Krise der Bank im Jahr 2001 nicht das zufällige Ergebnis unvorhersehbarer und überraschender Ereignisse war, sondern vielmehr einer sich ankündigenden langjährigen strategischen Fehlentwicklung des Unternehmens.

Auffällig ist, dass es sich sowohl bei den beiden hier dar-

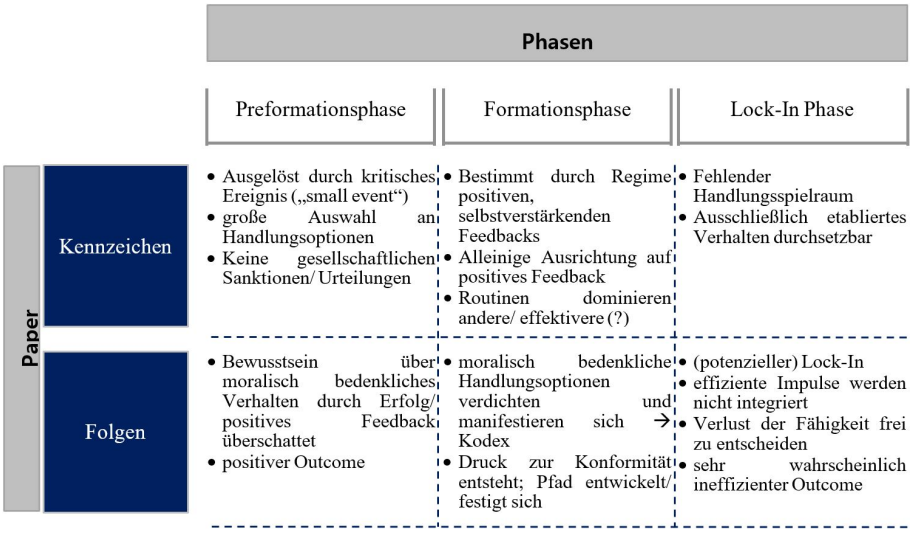

Abbildung 8: Kennzeichen und der einzelnen Phasen der Pfadabhängigkeit (Quelle: eigene Darstellung)

Abbildung 9: Integration der Forschungsfelder ”Unternehmenskrisen” und ”Pfadabhängigkeit von Unternehmen” (Quelle: eigene Darstellung)

gestellten als auch bei anderen Versuchen der Integration der Forschungsfelder fast ausschließlich um Einzelfallbetrachtungen handelt.

Einzelne Arbeiten aus dem Forschungsfeld der Pfadabhängigkeit bieten einen Anknüpfungspunkt für eine Integration: es kann davon ausgegangen werden, dass Unternehmen bzw. deren Mitglieder dazu bereit sind persistente, überholte Verhaltensweisen abzulegen – eine solche tiefgehende Veränderung kann u.a. durch Schocks oder Unternehmenskrisen herbeigeführt werden (vgl. Khalil, 2013: S. 27). Dieser Gedanke wird im Folgenden weiter ausgebaut. Abbildung 10 stellt eine weitere Darstellungsvariante der in Kapitel 2 im Abschnitt *Arten und Verlauf* vorgestellten Typisierung von Unternehmenskrisen dar. Sie zeigt den Verlauf einer Unternehmenskrise, gegliedert nach bedrohten Unternehmenspotentialen, im zeitlichen Verlauf. Neben Krisenart und Zeit als Dimension, wird in dieser Darstellung auch ein Handlungskorridor integriert, welcher sich von einem anfänglich großen zum letzten Endes kleinen Handlungsspielraum verändert. Aus der zuvor vorgestellten Literatur zu Unternehmenskrisen ist die Intuition folgender Abbildung 10 bekannt.

Diese Sicht auf die Entwicklung einer Strategischen Krise, der Art von hier im erweiterten Sinne untersuchter Krise, betont zum einen die Langfristigkeit der Krisenentwicklung und zum anderen den prozesshaften Automatismus („Dominostein-Prinzip“) einer initialen Fehlentscheidung.

Es wird offensichtlich, dass zu Beginn der Ereigniskette eine wie auch immer geartete Fehlentscheidung steht. Diese Entscheidung wird aus der uneingeschränkten Auswahl an Handlungsalternativen gewählt. Der destruktive Charakter der Krise entfaltet sich im Laufe der Zeit – zuerst unsichtbare, dann auch für die Öffentlichkeit sichtbare Krisenzeichen. Der Handlungsspielraum verengt sich bis hin zur Alternativlosigkeit. Bei Nichtreaktion auf jegliche Warnsignale mündet dieser Prozess im Zusammenbruch.

Dies war eine Betrachtung des Krisenprozesses aus Sicht der Forschung zu Unternehmenskrisen. Nun sollen die gewonnenen Erkenntnisse vor dem Hintergrund der Pfadabhängigkeitstheorie erörtert werden. Die Grundidee Organisationaler Pfadabhängigkeit nach Sydow et al. wurde bereits in Abschnitt 2.2.2 dargelegt. Die Autoren gehen neben der Beschreibung des Konstrukts der Pfadabhängigkeit auch auf ein Pfadbrechen ein. Das Pfadbrechen stellt dabei eine Möglichkeit dar, die es in einem pfadabhängigen Prozess gefangenen Unternehmen erlaubt, den Pfad zu verlassen. An dieser Stelle soll zuerst klar umrissen werden, was unter Brechen oder Lösen eines Pfads im organisationalen Kontext zu verstehen ist. Diese Entwicklung kann sowohl in der Komplexität als

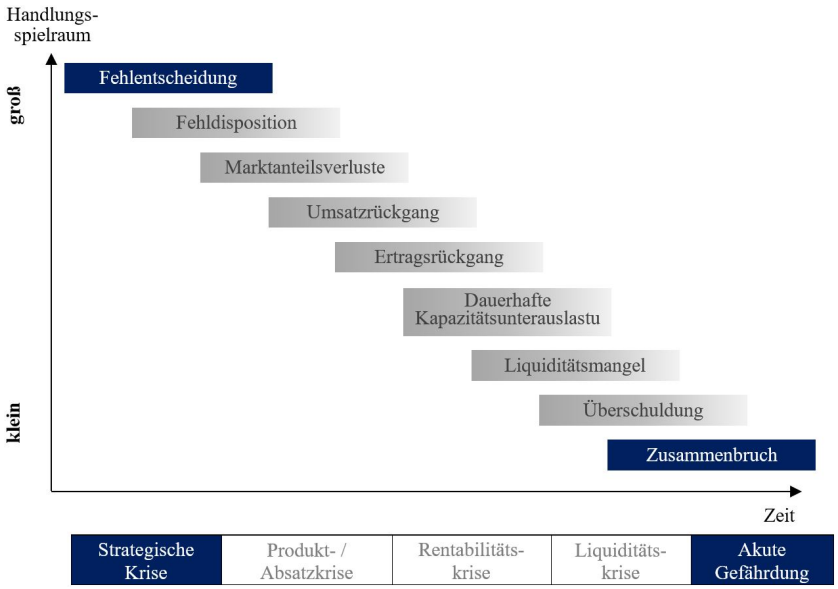

Abbildung 10: Verlauf einer Unternehmenskrise, beschrieben am Beispiel eines produzierenden Unternehmens (Quelle: In Anlehnung an Wöltje, 2011: S. 250)

auch in der Intensität stark variieren. Sydow und Koch definieren daher die Minimalbedingung des Brechens oder Lösens eines Pfads als Aktivitäten, welche „interrup[t] the logic and the specific energy of the self-reinforcing patterns of the process in question [as well as the] effective restoration of a choice situation“ (Sydow et al., 2009: S. 702). Ein Pfad ist folglich gelöst bzw. gebrochen, wenn eine zusätzliche, überlegene (Handlungs-) Alternative eine erneute Wahlmöglichkeit eröffnet.

Zwei Ansätze – der des bewussten Pfadbrechens und der des Pfadbrechens durch externe Kräfte – sollen diskutiert werden, wobei der Fokus hier auf dem letztgenannten liegt. Der erstgenannte Ansatz steht in gewisser Weise im Widerspruch zu sich selbst (vgl. Sydow et al., 2009: S. 702). Pfadabhängigkeit als eine Situation definiert, in welcher einzelne Mitglieder oder ganze Organisationen die Möglichkeit zwischen mehreren Alternativen zu wählen verloren haben, so ist die Unterstellung, dass eben diese Individuen den Pfad lösen können offensichtlich inkonsistent, denn „path-dependent behavior [...] excludes path-breaking behavior“ (Sydow et al., 2009: S. 702). Dennoch gilt der Grundsatz, dass kein Pfad ewig ist, und so beschreiben Sydow und Koch das Auflösen eines Pfads „through unforeseen exogenous forces, such as shocks, catastrophes, or crises” (Sydow et al., 2009: S: 701). Diese Ereignisse können das Unternehmen aufrütteln und ihm dabei helfen, sich aus der Enge der Pfadabhängigkeit zu lösen (vgl. Arthur, 1994: S. 118). Dieser Prozess des Pfadauflösens beruht in gewissem Maße auf dem Zusammenprall der internen Perspektive des Unternehmens, welche durch die Folgen der Pfadabhängigkeit geprägt sind, und der externen (gesellschaftlichen) Perspektive. Das Unternehmen ist in seinen eingefahrenen Strukturen und Verhaltensweisen gefangen. Aufgrund positiven Feedbacks wird auch in zunehmendem Maße eine Selbstreflektion der gewählten Strukturen und Verhaltensweisen unterbunden und selbst Impulse zur Selbstreflexion z.B. durch neue Mitarbeiter werden unterdrückt; das Unternehmen ist im Lock-In. Die Gesellschaft, in welche das Unternehmen eingebettet ist, entwickelt sich jedoch parallel weiter. Aufgrund unterschiedlicher Entwicklungen wächst die Diskrepanz zwischen interner und externer Perspektive zunehmen und kann sich mit voller Kraft entladen. Diese auf den ersten Blick negativ assoziierten Ereignisse können dem Unternehmen jedoch dabei helfen, die interne Perspektive an den Status Quo der externen Perspektive anzupassen und dabei den Pfad zu lösen. Folgende Abbildung 11 fasst die soeben beschriebenen Gedanken noch einmal bildlich zusammen:

Ein Pfad kann demnach verlassen bzw. gebrochen werden. Eine Situation nach dem Aufbrechen der Pfadabhängigkeit kann jedoch auch zum Ausbilden eines neuen Pfads bzw. einer neuen Pfadabhängigkeit führen (siehe Abbildung 12).

Ein Ausbrechen aus der Pfadabhängigkeit bzw. ein Wechsel auf einen anderen Pfad kann, durch einen exogenen Schock bspw. im Sinne einer Unternehmenskrise herbeigeführt werden. Der exogene Schock, ob in Form einer Krise oder einer anderen Katastrophe, rüttelt das System auf und gibt ihm die Möglichkeit, den Pfad zu verlassen bzw. zu wechseln. Dieser Prozess der Pfadauflösung kann evolutionär oder revolutionär erfolgen. Was bedeutet ein Pfadwechsel im hier erörterten Sinne? Innerhalb eines Unternehmens haben sich über mehrere Jahre hinweg z.B. korrupte Strukturen im Sinne eines Geschäftsverständnisses, dass die Zahlung von Schmiergeldern bei der Auftragsgewinnung als normal erachtet wird, manifestiert. Prallt eines Tages dieser festgefahrene Mechanismus (interne Perspektive) mit der sich ständig weiterentwickelnden externen Perspektive in Form einer Unternehmenskrise aufeinander, so muss das Unternehmen sich intern neu aufstellen. Es hat die Möglichkeit, den

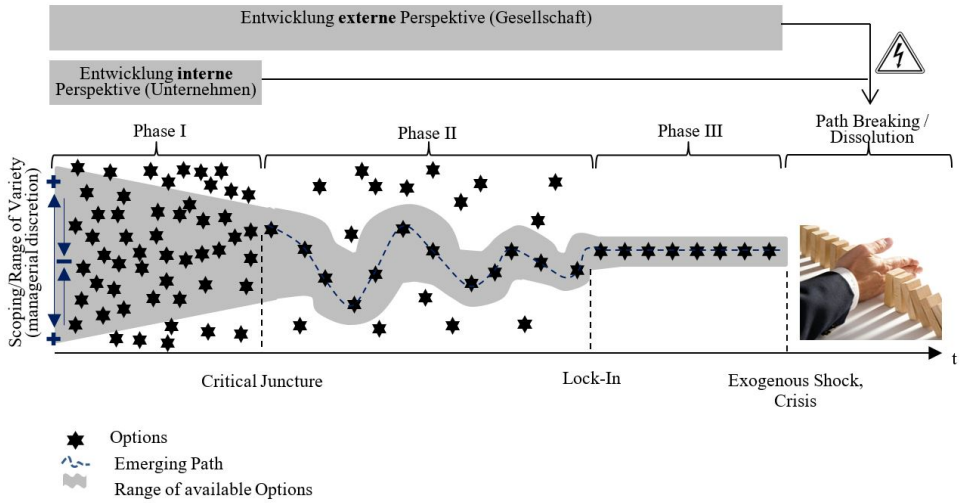

Abbildung 11: Krisen als logische Konsequenz von Pfadabhängigkeit? (Quelle: eigene Darstellung)

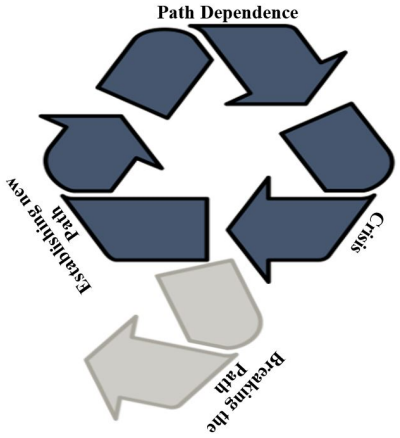

Abbildung 12: Kreislauf und Ausbrechen aus der Pfadabhängigkeit (Quelle: eigene Darstellung)

alten Pfad zu verlassen. Es besteht jedoch auch nach jeder Unternehmenskrise das Risiko, dass das Unternehmen einen neuen Pfad etabliert. Im erläuterten Fall würde dies bedeuten, dass das Unternehmen zwar keine Schmiergelder mehr zur Auftragsgewinnung einsetze, dafür aber z.B. Kinder-/ oder Schwarzarbeit, um weiterhin erfolgreich zu sein. Eine Unternehmenskrise hätte hier zwar den alten Pfad zerstört, jedoch wurde die Chance der Krise hier zugunsten eines anderen ethisch-moralisch negativ gearteten Pfads vertan. Eine solche Systembewegung kann des Weiteren auch durch einen schleichenden Wandel organisationaler Demographie sowie eine imperfekte Sozialisation neuer organisationaler Mitglieder ausgelöst werden (vgl. Sydow et al., 2009: S. 701). Dies soll jedoch nicht weiter vertieft werden.

Ziel dieses Abschnitts ist es, die Frage nach einer möglichen Vereinbarkeit von Unternehmenskrisen- und Pfadabhängigkeitsforschung zu beantworten. Obwohl selten gemeinsam untersucht, ermöglicht eine integrierte Betrachtung doch neue Erkenntnisse.

Abbildung 13 verdeutlicht, dass bei einer gemeinsamen Betrachtung (Verlauf einer Unternehmenskrise und Prozess der Pfadabhängigkeit) eine Fehlentscheidung am Anfang eines Krisen-Pfadabhängigkeitsprozesses stehen kann. Es wird aus einem uneingeschränkten Optionsspektrum eine Alternative gewählt. Am Ende dieser ersten Entwicklungsphase steht die critical juncture, die den Moment meint, welcher das Auftreten selbstverstärkender Effekte einläutet, die im weiteren Verlauf für die Persistenz verantwortlich sind und am Ende der zweiten Phase zum Lock-In führen. Ein in Phase 3 befindliches Unternehmen verfügt über keine Handlungsalternativen mehr, es ist in der historischen Entscheidung gefangen. Kommt es nun jedoch bedingt durch den Zusammenprall interner und externer Perspektiven zu einer Unternehmenskrise, so erleidet das Unternehmen zwar eine Krise, jedoch ist diese gleichzeitig eine Chance, mit dem Pfad zu brechen. Der Entscheidungsspielraum erweitert sich; weitere, überlegene Handlungsalternativen stehen zur Auswahl.

Diese gemeinsame Betrachtung legt die Vermutung nahe, dass pfadabhängige Unternehmen mit relativ hoher Wahrscheinlichkeit in ihrem Entwicklungsprozess eine Krise als

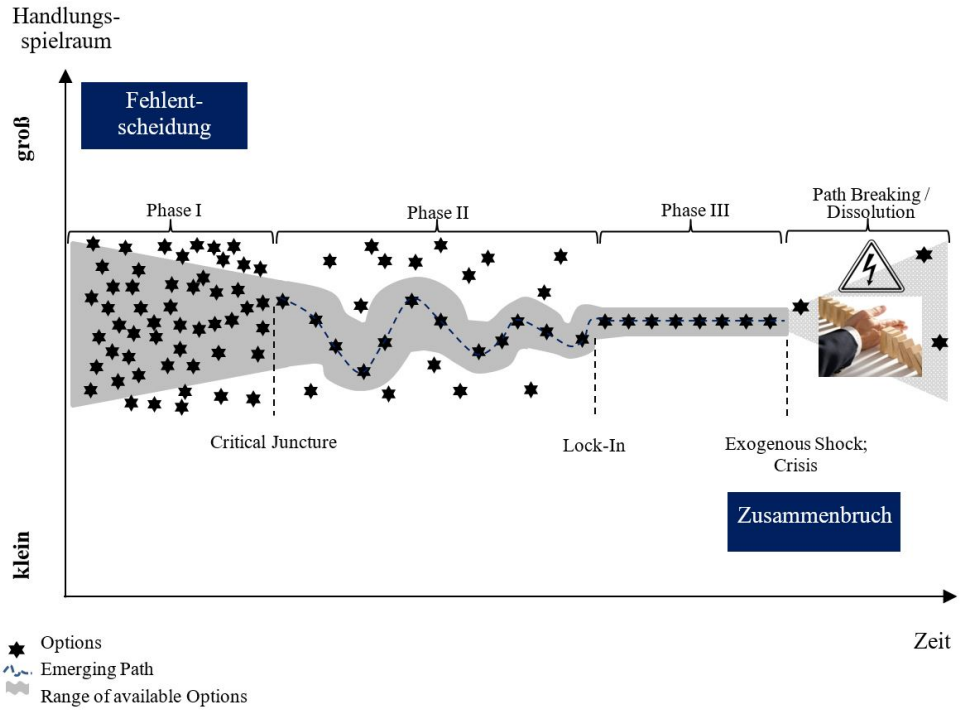

Abbildung 13: Vereinbarkeit von Unternehmenskrisen- und Pfadabhängigkeitsforschung (Quelle: eigene Darstellung)

pfadbrechendes Ereignis erleiden, denn „no path is forever“ (Sydow et al., 2009: S. 701). Vergleicht man also pfadabhängige Unternehmen mit pfadunabhängigen Unternehmen, so würde angenommen werden, dass bei pfadabhängigen Unternehmen relativ mehr Unternehmenskrisen zu beobachten sind. Dies gilt es zu einem späteren Zeitpunkt zu untersuchen. Eine weitere interessante Frage ist die nach möglichen Ansatzpunkten zur Prävention von Unternehmenskrisen. Wenn Unternehmenskrisen als logische Konsequenz von Pfadabhängigkeitsprozessen auftreten, so sollten vermehrte Forschungsanstrengungen zur Schaffung von Strukturen unternommen werden, welche eine verzögerte Parallelentwicklung der internen Perspektive (Unternehmen) vermeiden.

3. Fallbasierte Untersuchung relevanter ethisch-moralisch bedingter Unternehmenskrisen

Dieser Abschnitt widmet sich der Methodik und dem Vorgehen dieser Arbeit. Einführend wird die Methodik der fallbasierten Untersuchung in ihrem theoretischen Zusammenhang erläutert. Anschließend wird das konkrete Vorgehen zur Identifikation und Kategorisierung ethisch-moralisch bedingter relevanter Krisenfälle beschrieben. In den Abschnitten 3.2 und 3.3 sind die Ergebnisse der fallbasierten Untersuchung von Deutschland und den USA zu finden.

3.1. Methode der fallbasierten Untersuchung

Die Fallstudienforschung ist eine mittlerweile etablierte Forschungsmethode qualitativer Sozialwissenschaften und wurde zunächst vor allem zur Erforschung sozialer Phänomene in der Medizin und den Rechtswissenschaften eingesetzt (vgl. Tellis, 1997: S. 2). Die qualitative Forschung in Abgrenzung zur quantitativen Forschung eignet sich im Besonderen dazu wenig untersuchte Forschungsgebiete zu analysieren (vgl. Asprion, 2013: S. 183, Eisenhardt, 1989: S. 532). Lange Zeit galten qualitativ orientierte Forschungsmethoden als zweitbeste Wahl gegenüber quantitativen Methoden. Dabei haben qualitative Forschungsmethoden allgemein bzw. Fallstudien im Speziellen auf internationalem Terrain in den vergangenen Jahrzehnten großen Zulauf erfahren (vgl. Borchardt und Göthlich, 2007: S. 34). Yin definiert eine Fallanalyse als „an empirical inquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident [...]“ (Yin, 1984: S. 23). Eine fallbasierte Untersuchung kann als Prozess betrachtet werden, der einerseits iterativ ist und sich andererseits stark an Daten orientiert (vgl. Eisenhardt, 1989: S. 532) und dabei eine „möglichst ganzheitliche Beschreibung der Situation und des Problems [...] einer Institution“ (Universität Hamburg, 2014) liefert. Im Mittelpunkt steht die Ermittlung der Faktoren, die maßgeblichen Einfluss auf die Situation ausüben. Die daraus gewonnene Erkenntnis soll in einem nächsten Betrachtungsschritt Aufschluss über mögliche effektive Krisenpräventionsmaßnahmen geben. Darüber hinaus kann eine fallbasierte Untersuchung auch das Aufzeigen möglicher Handlungsalternativen beinhalten. Diese Art der empirischen Arbeit spiegelt meist eine bestimmte Perspektive wider. Demzufolge ist kein vollständiges Abbild der Wirklichkeit zu erwarten, sondern eine „bewusste Aus-

wahl aus der Fülle möglicher Daten der Lebenswirklichkeit“ (Universität Hamburg, 2014).

3.1.1. Vorgehensweise

Ein Fall ist im Verständnis dieser Arbeit eine ethisch-moralisch bedingte Unternehmenskrise nach den in Abschnitt 2.1.2 beschriebenen Merkmalen. In einem ersten Schritt werden relevante Krisenfälle mittels einer dokumentierten Suchstrategie identifiziert. Anschließend wird eine Einteilung der identifizierten Fälle in drei definierte Kategorien vorgenommen. Der Ergebnisdarstellung, unterteilt nach den beiden untersuchten Ländern Deutschland und USA, folgt eine Bewertung der Ergebnisse (siehe auch 14).

3.1.2. Identifikation und Selektion zu analysierender Fälle

Der Suche nach ethisch-moralisch bedingten Unternehmenskrisen wird der Zeitraum seit der Einführung des DAX im Juni 1988 als pragmatischer Startzeitpunkt zugrunde gelegt. In die Untersuchung einbezogen werden folglich Fälle, die sich zwischen Juni 1988 und Mai 2014 ereignet haben.

Die ausschließliche Inklusion von DAX 30 bzw. Dow Jones 30-Unternehmen soll eine Vergleichbarkeit der untersuchten Unternehmen bzw. Fälle in Bezug auf Internationalität, Umsatz etc. annähern und Drittfaktoren soweit möglich konstant halten.

Die Auswahl des recht langen Zeitraums lässt sich damit begründen, dass Unternehmenskrisen mit relativ großer zeitlicher Unregelmäßigkeit auftreten und somit ein breit gefasster Zeitrahmen nötig ist, um ausreichend Fallmaterial als Grundlage für die fallbasierte Untersuchung sicherzustellen. Um eine ländervergleichende Betrachtung deutscher und amerikanischer Ergebnisse durchführen zu können, werden einerseits Unternehmen des DAX 30 (umsatzstärkste und gewichtigste an der Frankfurter Börse registrierte Unternehmen) und andererseits des Dow Jones 30 (größte amerikanische Standardwerte) analysiert. Untersucht werden alle Unternehmen, die innerhalb des betrachteten Zeitraums je unter den DAX 30 bzw. Dow Jones 30-Unternehmen waren oder/ und es aus bestimmten Gründen nicht mehr sind. Bei der Identifikation der Fälle werden bewusst Sekundärdaten (Pressemitteilungen, Zeitschriftenartikel, Geschäftsberichte etc.) als Informationsgrundlage herangezogen, da

- Primärdaten z.B. durch Befragung von Unternehmensrepräsentanten für solch kritische Themen sehr schwierig zu erhalten sind bzw. durch die Unternehmensperspektive verzerrt sein könnten,
- Sekundärdaten eine größere Vielfalt aufweisen und damit eine abstrahierendere Betrachtung und Evaluation der Situation erlauben und
- weitaus leichter verfügbar sind, berücksichtigend, dass ein durchaus langer Zeitraum zu untersuchen ist.

Im Online-Appendix befinden sich Listen aller Unternehmen, die seit 1988 pro Jahr im DAX 30 bzw. Dow Jones 30 gelistet waren.

Bzgl. der Suche nach relevanten Fällen wurde in drei Such- bzw. Selektionsschritten vorgegangen, über welche die folgende Abbildung 15 eine Übersicht liefern soll. Die dokumentierten Teilergebnisse der einzelnen Suchschritte finden sich in der beigelegten Excel-Tabelle.

3.1.3. Kategorisierung identifizierter Fälle

Zur späteren Evaluierung effektiver Präventionsmechanismen sowie zur Überprüfung des Einflusses der Pfadabhängigkeit soll eine zweidimensionale Matrix herangezogen werden. Im Folgenden werden die beiden Dimensionen der Maßnahmen sowie der Pfadabhängigkeitsmatrix in Abbildung 16 vorgestellt.

1) Kategorien von Unternehmen Diese Dimension gliedert die identifizierten Fälle nach der Anzahl ihrer Krisen unter den „Megakrisen“, d.h. den größten ethisch-moralisch bedingten Unternehmenskrisen im betrachteten Zeitraum. In Kategorie 1 finden sich alle Unternehmen, für welche mindestens zwei Megakrisen zu finden waren. Kategorie 2 fasst alle Unternehmen zusammen, welche eine Megakrise aufzuweisen haben. Kategorie 3 beinhaltet als Residualgruppe alle Unternehmen, welche keine Megakrise durchlebt haben. Zusätzlich wurde innerhalb jeder Kategorie jedes Unternehmen auf das weitere Vorkommen ethisch-moralisch bedingter Krisen, welche nicht zu den Megakrisen zählen, untersucht. Das Forschungsinteresse dieser zusätzlichen Betrachtung gilt dem Aufdecken und der Analyse möglicher zeitlicher Ablaufschemata verschieden stark ausgeprägter Krisenfälle. Von der auf Einzelkrisen bezogenen Betrachtung wurde anschließend auf eine Analyse auf Unternehmensebene übergegangen. Grund hierfür ist, dass sich eine rein zeitliche Beurteilung der Einführung bestimmter Maßnahmen aufgrund mangelnder Berichterstattung und verfügbarer Daten sehr schwierig gestaltete. Statt „die Siemens AG hat als Reaktion auf Krise Siemens2 Maßnahme XY ergriffen“ (siehe Kodierung in der beigefügten Excel-Tabelle) wird an dieser Stelle eine Bestandsaufnahme im Sinne „die Siemens AG führt Maßnahme XY durch“ erörtert. Eine Beurteilung der Kausalität einzelner Unternehmenskrisen für die Einführung einer Maßnahme war aufgrund der Informations- und Berichterstattungslage nicht möglich bzw. hätte ein zu hohes Maß an Subjektivität involviert. Dieser Gefahr wird durch die reine Bestandsaufnahme an Maßnahmen je Unternehmen vorgebeugt. Um eben diese Betrachtung auf Unternehmensebene durchführen zu können, wurden alle zum Dekaden-Median relativen Google-Hits Werte einzelner Unternehmenskrisen eines Unternehmens addiert und in der Größe absteigend geordnet (siehe Berechnungen in beigefügter Excel-Tabelle). Diesem Schritt folgt eine binäre Trennung der so geordneten Unternehmen in krisenanfällige und krisenunanfällige Unternehmen. Hierzu wurde die Regel festgelegt, dass alle Unternehmen mit einem relativen Google-Hits Wert gleich bzw. größer 1000 Prozent als krisenanfällig und solche mit einem relativen Google-Hits Wert kleiner 1000 Prozent als krisenunanfällig betrachtet werden. Mithilfe dieser Kategorisierung werden sowohl potenzielle Krisenpräventionsmaßnahmen als auch Kriterien der Pfadabhängigkeit beurteilt.

Abbildung 14: Ablaufschema der fallbasierten Analyse ethisch-moralisch bedingter Unternehmenskrisen (Quelle: eigene Darstellung)

2) Potenzielle Präventionsmaßnahmen Um zu einem späteren Zeitpunkt in dieser Analysephase eine Aussage über potenziell effektive Präventionsmechanismen machen zu können, wurde eine Übersicht solcher erarbeitet. Anspruch auf Vollständigkeit kann an dieser Stelle aufgrund der Vielfalt an Maßnahmen nicht erhoben werden, was die Interpretation der Ergebnisse jedoch nicht schmälert. Die im Folgenden dargestellte Übersicht möglicher Präventionsmaßnahmen wurde zum einen deduktiv, d.h. aus der Literatur zu Unternehmenskrisen, abgeleitet, was einen eher theoretisch orientiertem Ansatz entspricht. Zum anderen wurde induktiv vorgegangen, d.h. aus dem Material heraus, was einer eher praxisorientierten Vorgehensweise entspricht. Im Laufe des Prozesses der Fallidentifikation gefundene Mechanismen bzw. Maßnahmen wurden der Liste hinzugefügt. Dieser symbiotische Ansatz spiegelt den Charakter des hier diskutierten Themas wider: Theorie und Praxis treffen aufeinander und ergänzen sich (im Idealfall) gegenseitig. Die Eignung diese Maßnahmen zur Prävention ethisch-moralisch bedingter Unternehmenskrisen soll im Rahmen dieser Arbeit untersucht werden. Den in Tabelle 2 vorgestellten Maßnahmen wird eine krisenpräventive Wirkung unterstellt, d.h. ihr Einsatz soll zur Krisenfreiheit eines Unternehmens beitragen. Dieser Funktion entsprechend wird als Ergebnis des durchzuführenden χ^2-Unabhängigkeitstest eine signifikante negative Korrelation der „Krisenanfälligkeit von Unternehmen“ und den elf Maßnahmen erwartet.

Bezüglich der Pfadabhängigkeit dient folgende Matrix in Abbildung 17 als Grundlage der Betrachtung und Auswertung.

3) Kriterien der Pfadabhängigkeit Um die identifizierten krisenanfälligen sowie –unanfälligen Unternehmen auf Anzeichen einer Pfadabhängigkeit hin zu untersuchen, wurden drei Kriterien als Kennzeichen einer pfadabhängigen Entwicklung definiert.

Persistenz: Wie bereits dargelegt, handelt es sich bei der Pfadabhängigkeit um ein langfristiges Konzept. Bezogen auf Unternehmen bzw. Unternehmenskrisen kann daraus geschlossen werden, dass (ethisch-moralisch unsaubere) Strukturen oder Verhaltensweisen sich über einen langen Zeitraum ausbilden und erst dann als festgefahren oder eben persistent zu beschreiben sind. Untersucht wird daher je Krisenfall, ob als Krisenursache geltende Strukturen oder Verhaltensweisen mindestens über zwei Jahre ausgeübt wurden. Erst ein langzeitiges und starres Ausüben bestimmter Strukturen oder Verhaltensweisen wird als Hinweis auf eine mögliche Pfadabhängigkeit gewertet.

Ausmaß: Pfadabhängigkeit bezeichnet des Weiteren Strukturen oder Verhaltensmuster, welche von Gruppen oder gar dem ganzen Unternehmen hingenommen bzw. ausgeübt werden. Ein dominantes Schema trägt sich nur über die Adaption mehrerer Personen weiter. Ein weiteres hier untersuchtes Kennzeichen ist somit das Ausmaß involvierter Personen, welche die Struktur bzw. das bestimmte Verhaltensmuster leben. Ein großes Ausmaß, d.h. ein großer involvierter Personenkreis, deutet auf Pfadabhängigkeit hin, wenn in der Berichterstattung nicht ausdrücklich erwähnt wird, dass nur einzelne Personen dieses gewisse ethisch-moralische Fehlverhaltens ausführen.

Themenkonzentration: Ein Pfad steht in der Theorie Organisationaler Pfadabhängigkeit, wie bereits erläutert, für eine gewisse festgefahrene, persistente Struktur bzw. ein Verhaltensmuster. Sind bei Unternehmen über den betrachteten Zeitraum immer wieder dieselben Krisenthemen wie bspw. Spionage, Korruption oder Betrug präsent, so wird dies als Indiz für eine mögliche Pfadabhängigkeit gewertet.

3.1.4. Bewertung mithilfe eines χ^2-Unabängigkeitstests

Für die anschließende Beurteilung der kategorisierten Ergebnisse wird ein Pearson χ^2-Unanhängigkeitstests durchge-

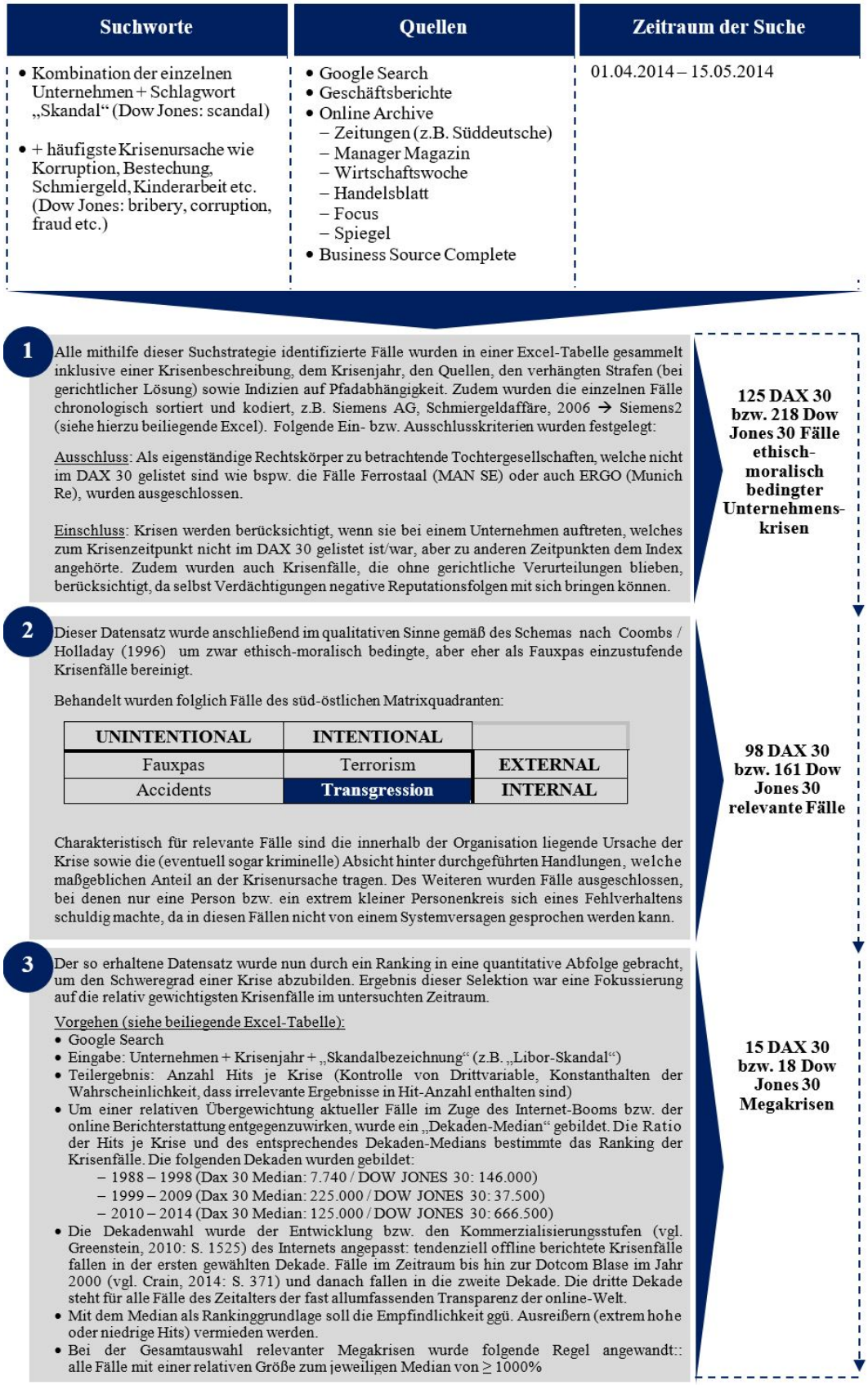

Abbildung 15: Beschreibung der Suchstrategie nach relevanten Fällen sowie Ergebnisse (Quelle: eigene Darstellung)

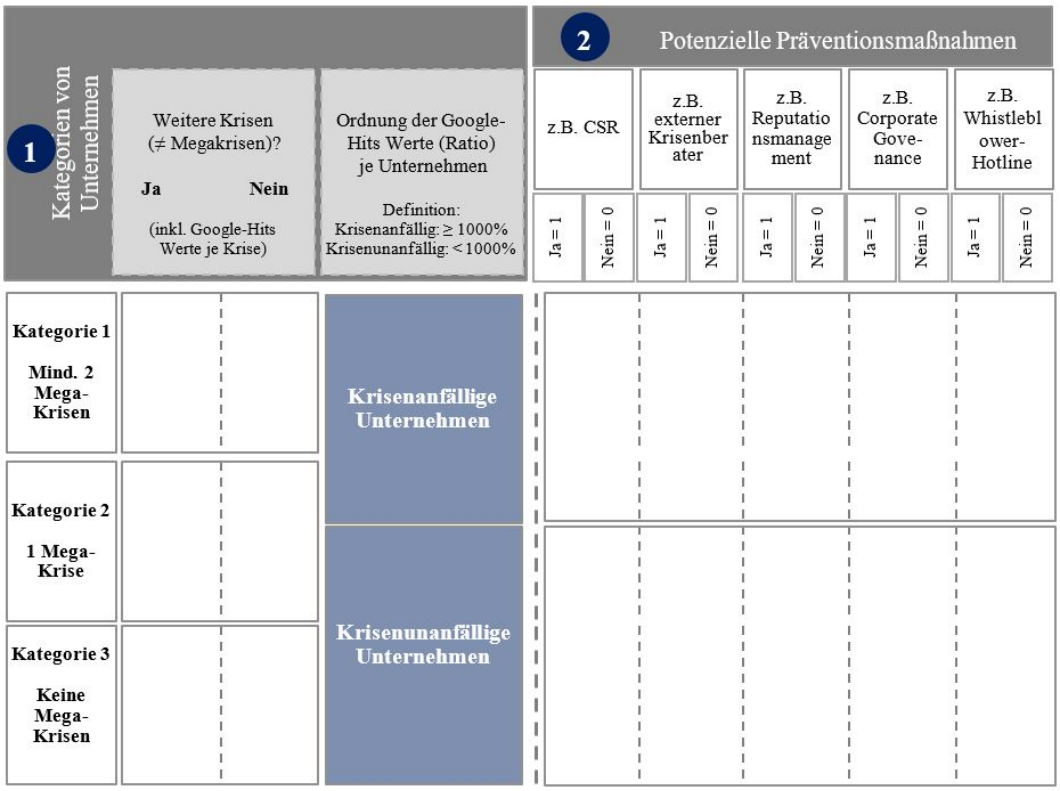

Abbildung 16: Kategorisierung identifizierter Fälle nach zwei Dimensionen (potenzielle Präventionsmechanismen) (Quelle: eigene Darstellung)

führt. Dieser statistische Test wird dabei sowohl zur Überprüfung der These der Unabhängigkeit bezogen auf die potenziellen Krisenpräventionsmaßnahmen als auch bezüglich der Kriterien der Pfadabhängigkeit herangezogen. Ziel des Tests ist in beiden Fällen zu verifizieren, ob eine Korrelation zwischen zwei nominal skalierten Variablen besteht bzw. das Verhältnis der Variablen zueinander zu interpretieren. Die zugrunde liegenden Hypothesen sind:

> H_0: Variable X und Y sind unabhängig voneinander
>
> H_1: X und Y sind nicht unabhängig voneinander

Die Berechnungen beziehen sich auf Vierfeldertafeln, wodurch sich ein Freiheitsgrad von eins und ein entsprechender kritischer χ^2-Wert zu einem bestimmten Siginifikanzniveau ergeben. Ho, also die Unabhängigkeitsvermutung, kann nur abgelehnt werden, wenn der entsprechende χ^2-Wert größer als der entsprechende kritische Wert ist. Folglich lässt ein hoher χ^2-Wert und damit eine Ablehnung der Nullhypothese auf eine Korrelation der beiden Merkmale schließen (vgl. Toutenburg und Heumann, 2000: S. 272). Diese Prüfstatistik lässt jedoch keine Beurteilung der Stärke der Korrelation zu. Hierfür werden ergänzende Korrelationsmaße hinzugezogen:

- Phi-Koeffizient (normierter Wertebereich: -1 bis +1)
- Kontingenzkoeffizient C (normierter Wertebereich: 0 bis +1)
- Cramer's V (normierter Wertebereich: 0 bis +1)

Da im vorliegenden Fall Vierfeldertafeln behandelt werden, sind die Maße Cramer's V und Phi dem Betrag nach identisch (vgl. Schwaiger, 2013: S. 5-8). Nähere Erläuterungen und Interpretationen der einzelnen (Teil-) Ergebnisse finden sich in den nächsten beiden Abschnitten (Deutschland, USA).

3.2. Deutschland (DAX 30)

In den folgenden sechs Abschnitten werden die erarbeiteten Ergebnisse für die DAX 30-Unternehmen dargestellt. Eine Bewertung bzw. ein Vergleich mit den Dow Jones 30-Ergebnissen findet im anschließenden Bewertungsabschnitt 3.4 statt.

3.2.1. Absolute Anzahl ethisch-moralisch bedingter Unternehmenskrisen

Derzeit zählt der Krisennavigator des Kieler Instituts für Krisenforschung rund 270 Fälle pro Jahr. Statistisch betrachtet heißt das: Alle eineinhalb Tage entsteht bei einem Unternehmen in Deutschland ein Skandal. In Deutschland wie auch im internationalen Umfeld, verändert sich die Krisenlandschaft zunehmend. Krisenauslöser sind weniger operative Fehler, sondern die Entwicklung zeigt eine starke Zunahme kommunikativer Krisenfälle, wie folgende Abbildung 18 veranschaulicht.

Tabelle 2: Übersicht untersuchter potenzieller Krisenpräventionsmaßnahmen

(Eigene Darstellung).

Maßnahme	Quelle/ Autoren
Externe, unabhängiger Krisenberater	Unabhängige Berater oder externe Aufseher (wie im Daimler-Fall Ex-FBI Chef Louis Freeh) von außerhalb des betroffenen Unternehmens, um Vertrauen zu schaffen und die Ernsthaftigkeit der Besserungsbemühen glaubhafter zu machen. Diese unterstützen die Umsetzung (gerichtlich) angeordneter Maßnahmen, schaffen „saubere" Strukturen oder liefern Best-Practice Ansätze mit dem Ziel der künftigen Krisenfreiheit (vgl. z.B. Polley und Kroner, 2012: S. 69).
Corporate Compliance-System/ Programm · Code of Conduct vorhanden, · Offenlegung von Compliance Regeln etc. (vgl. McGreal, 2013: S. 108)	Compliance wird definiert als „conforming to a rule, such as a policy framework, standard or law" (Prorokowski und Prorokowski, 2014: S. 65). In der Praxis wird Corporate Compliance z.B. mit überprüfbaren Kriterien für den Umgang mit Geschäftspartnern, mit eindeutigen Regeln für die Auftragsakquise, mit strengen Kontrollen und Konsequenzen bei Verfehlungen verbunden (vgl. z.B. Heybrock, 2012: S. 10). Unter Compliance werden demzufolge eine Vielzahl von Einzelmaßnahmen gebündelt, die das rechtmäßige Verhalten eines Unternehmens, der Mitarbeiter sowie der Leitungs- und Aufsichtsorgane sichern soll (vgl. Bayrak et al., 2012: S. 8) und dabei vor allem den Aspekt der Vorbeugung von Straftaten verfolgt (vgl. Bussmann et al., 2010: S. 11).
Whistleblower-Hotline	Anonymes Hinweisgebersystem bzw. Hotline, über welche Mitarbeiter die Möglichkeit haben, anonym Fälle verdeckter Wirtschaftskriminalität oder anderer ethisch-moralischer Auffälligkeiten im Unternehmen zu melden. Hauptkennzeichen sind dabei z.B.: · Tip Input: Hinweise werden über viele Kanäle angenommen (E-Mail, Telefon etc.) · Anonymität: vertrauliche Behandlung ist zu jedem Zeitpunkt garantiert · Erreichbarkeit: 365/24/7-Grundsatz · Mehrsprachigkeit: mehrsprachige Ansprache und Auskunft möglich etc. (vgl. z.B. Steinbuch, 2013; Andrews und LeBlanc, 2013: S. 33)
Red-Flags System	Risiko- oder Schwachstellenanalyse, durch welche besonders gefährdete Unternehmensbereiche für Korruption oder Betrug identifiziert, gekennzeichnet und Maßnahmen gegen diese eingeleitet werden (vgl. z.B. Landeskriminalamt NRW, 2007: S. 7).
Corporate Social Responsibility (CSR)	CSR bezeichnet „discretionary business practices and contributions of corporate resources intended to improve societal well-being (Kotler und Lee, 2005: S. 3). CSR transportiert die der Unternehmung zugrundeliegenden Werte, was dazu beitragen kann, dass (potenzielle) Stakeholder eine starke Bindung mit dem Unternehmen aufbauen. Das Unternehmen kann von dadurch ausgelöstem, begünstigtem Verhalten profitieren (vgl. Korschun et al., 2014: S. 23).
Rotationsprinzip	Rotation meint den regelmäßigen Wechsel von Positionen, vor allem in moralisch verantwortungsvollen Positionen. Durch eben diese Rotation steigen die Opportunitätskosten betrügerischer oder korrupter Handlungen. Die vermeintlichen Betrugspartner müssen sich immer wieder mit neuen Partnern arrangieren. Der Aufbau langfristiger Beziehungen wird hierdurch gestört. Zudem erhöht sich das Entdeckungsrisiko (vgl. Schmidt, 2004: S. 23).

(Continued)

Tabelle 2—continued

Maßnahme	Quelle/ Autoren
Reputations-management	Die Unternehmensreputation besteht aus „a set of key characteristics attributed to the firm by various stakeholders" (Bromley, 1993). Diese Charakteristika beziehen sich vor allem, aber nicht ausschließlich, auf Attribute des Qualitätsmanagements, die Verantwortungsübernahme ggü. der Gesellschaft, die Qualität der Produkte bzw. Dienstleitungen etc. (vgl. Fombrun und Shanley, 1990: S. 233). Managementaktivitäten, welche das Ziel verfolgen, eben diese Wahrnehmung des Unternehmens der Stakeholder zu beeinflussen, zählen als Reputationsmanagement (vgl. Carter, 2006: S. 1146).
Issues Management	Bereits Henry Kissinger, ehemaliger US-Außenminister, erkannte, dass „an issue ignored is a crisis invited" - Ein Thema zu ignorieren, ist eine Garantie für eine Krise. Issues Management bezeichnet in Anlehnung an diese Einsicht, ein Frühwarn- und Reaktionssystem, welches das Unternehmensumfeld beobachtet und analysiert und so Krisenprävention aber auch Chancenkommunikation ermöglicht (vgl. Heath und Palenchar, 2008: S. 9, Fleck, 2014).
Risiko-/ Chancen-management	Unter Risikomanagement wird die Messung und Steuerung aller betriebswirtschaftlichen Risiken und Chancen unternehmensweit verstanden (vgl. Wolke, 2008: S. 1).
Corporate Governance Strukturen + Reporting	Corporate Governance regelt Rechte, Aufgaben und Verantwortlichkeiten organisationaler Organe (Geschäftsführung, Aufsichtsrat und Vorstand), Mitarbeiter sowie weiteren Stakeholdern. Es wird das Ziel verfolgt, einen Interessensausgleich der beteiligten Stakeholder zu erreichen (vgl. Wolf, 2003: S. 2).
Integrity Management	Integrität im organisationalen Kontext kann verstanden werden als „a philosophy of consistency of actions, values, methods, organizational principles, expectations, and results" (Odrakiewicz und Odrakiewicz, 2014: S. 65). Managementaktivitäten, welche sich der Umsetzung dieser Philosophie widmen, werden als Integrity Management bezeichnet.

Von allgemeinen deutschen Krisenfällen zu den hier relevanten übergehend bietet der Online-Appendix eine Übersicht aller Unternehmen, die im Zeitraum von 1988 bis 2014 zum DAX 30 zähl(t)en. Diese stellen die Grundgesamtheit der Untersuchung für den deutschen Markt dar.

Die folgende Matrix in Abbildung 19 zeigt das Ergebnis des soeben vorgestellten Identifikations- sowie Selektionsprozesses relevanter Krisenfälle für Deutschland.

Unabhängig davon, welche Unternehmen hinter den hier dargestellten Krisenfällen stehen, erlaubt die Übersicht folgende Einsichten:

- Nur drei Unternehmen der untersuchten DAX 30-Unternehmen fallen in Kategorie 1. Diese drei Unternehmen sind die Deutsche Post AG, die Siemens AG sowie die Volkswagen AG, also durchaus sehr große Konzerne mit einer starken Öffentlichkeitspräsenz sowie endkundenrelevanten Produkten bzw. Dienstleistungen.

- Festzustellen ist, dass die Kategorie 1-Unternehmen allesamt neben den mindestens zwei Megakrisen noch weitere, kleinere Krisen aufzuweisen haben. Diese Tatsache kann ein Indiz auf Pfadabhängigkeitsentwicklungen sein – nach dem frei gewählten Leitsatz „große Krisen kommen nie alleine". Jedoch ist hier neben der Feststellung, dass es weitere, nicht den Megakrisen zugeordnete Krisen gibt, auch die Chronologie bzw. das Muster der Krisen zu berücksichtigen. Diese systematische Betrachtung folgt in Abschnitt 3.2.4.

- Im Unterschied zu Kategorie 1-Unternehmen existieren in Kategorie 2 auch Unternehmen, welche zwar eine Megakrise, aber keine weiteren relevanten Krisen erlebt haben (MAN und SAP). Diese Feststellung zeigt jedoch, dass es durchaus für große Unternehmen möglich ist, mit einer eventuellen Krisenvergangenheit abzuschließen und fortan krisenfrei zu wirtschaften.

- Unter den Kategorie 3-Unternehmen ist besonders die Gruppe der elf Unternehmen interessant, die weder einer Megakrise noch sonstige, kleinere ethisch-moralisch bedingte Krisen erlebt haben. Deren nähere Betrachtung folgt in Abschnitt 3.2.3.

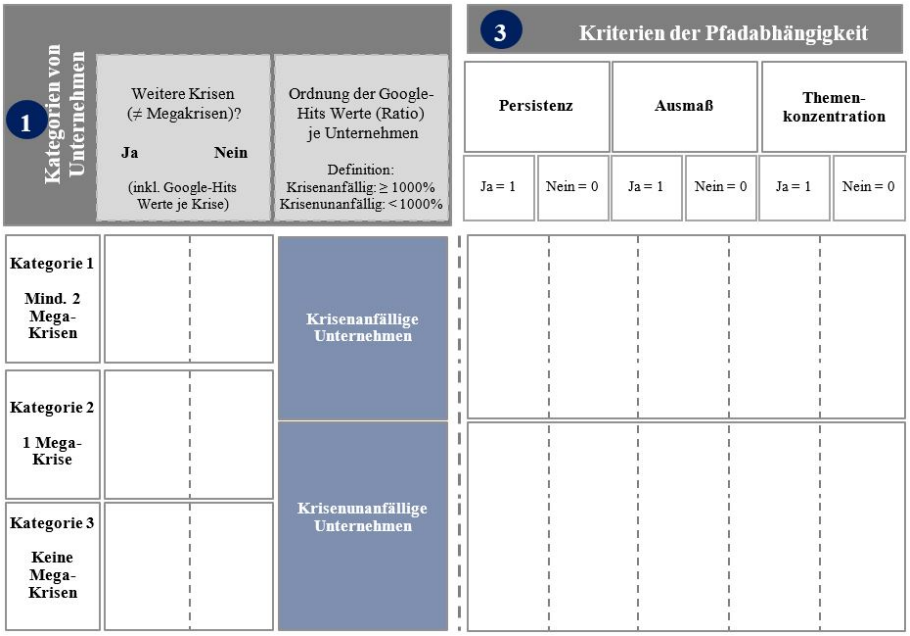

Abbildung 17: Kategorisierung identifizierter Fälle nach zwei Dimensionen, Kriterien der Pfadabhängigkeit (Quelle: eigene Darstellung)

3.2.2. DAX 30 „Megakrisen“

Um nun von der „Makro-Betrachtung“ (Anzahl der Unternehmen je Kategorie) in die „Mikro-Betrachtung“ (Details zu einzelnen Fällen) überzugehen, soll folgende Tabelle 3 eine Übersicht geben. Dargestellt sind die 15 identifizierten DAX 30-Megakrisen sortiert nach ihrer Ratio (Anzahl der Google Hits je Krise im Verhältnis zum spezifischen Dekaden-Median) sowie das jeweilige Krisenjahr und das Krisenthema. Die Kodierung der einzelnen Krisen (z.B. „MAN1“) ist in der beiliegenden Excel-Tabelle nachzulesen. Die farbliche Kennzeichnung referenziert auf das zuvor eingeführte Kategorienschema und bezeichnet, wie in der Legende dargestellt, die Kategorien 1 bis 3.

MAN führt mit der 2009 aufgedeckten Korruptions-Krise die Liste der Megakrisen an. Interessant an dieser Stelle zu beobachten ist, dass MAN zwar die größte Krise nach dieser Liste aufweist, aber keine weiteren Krisen zu verzeichnen hat. Es zeigt sich zudem, dass vor allem die „Großen der Großen“ in diesem Ranking auftauchen, d.h. die wohl größten Konzerne des Index‘. Diese erleben offensichtlich sowohl aufgrund positiver Berichterstattung über z.B. wachsende Umsatz- oder Beschäftigungszahlen als auch wegen negativer Meldungen wie im Fall ethisch-moralisch bedingter Krisen großes Aufsehen.

Auffällig ist des Weiteren, dass in acht von 15 Krisenfällen Korruption im Mittelpunkt der Krise steht. Bestechung und Spionage sind die nächstfolgenden Krisenthemen. Da Korruptionsfälle hier eine wichtige Rolle spielen, soll ein kurzer Blick auf politische bzw. gesellschaftliche Hintergründe geworfen werden. Korruption als gesamtgesellschaftliches Phänomen zieht sich durch die unterschiedlichsten Lebensbereiche – von der Politik über das Gesundheitswesen bis hin zur freien Wirtschaft. Die unabhängige Organisation Transparency International Deutschland e.V. definiert Korruption als die Situation, in der „jemand die ihm anvertraute Macht zum privaten Vorteil missbraucht“ (vgl. Transparency, 2014). Laut Bundeskriminalamt kennen sich die „Täter“ in rund 90 Prozent aller Korruptionsfälle über mehrere Jahre – daher oft die Bezeichnung als Strukturelle Korruption (vgl. Altmeier, 2012). Dass Korruption als Delikt aufgefasst wird, ist allerdings keine historische Selbstverständlichkeit. Bereits 1977 wurde der Foreign Corrupt Practices Act (FCPA) als Bundesgesetz in den USA erlassen. Dem Gesetz folgend sind Zuwendungen an (ausländische) Amtsträger mit dem Zweck den Zuschlag für ein Geschäft zu erhalten oder eine Geschäftsbeziehung aufrechtzuerhalten, verboten. Das Verbot gilt für alle Zahlungen oder Angebote von Geld zum Zweck der Korruption (vgl. Bayrak et al., 2012: S. 18). Ein entsprechendes Verbot besteht in Deutschland erst seit 1999. Bis 1999 war Korruption in Deutschland nur moralisch anstößig (vgl. Leyendecker, 2011), wurde aber strafrechtlich nicht verfolgt. Der Grundsatz der ”Zero Tolerance” gilt erst seitdem (vgl. Freitag und Katzensteiner, 2012). In den vergangenen Jahren ist in Kombination mit der gesetzlichen Strafbarkeit die Aufklärungsrate gestiegen. Nicht, weil es zwangsläufig mehr Korruption gibt, sondern weil die Gesellschaft kritischer geworden ist. In nicht allzu weit entfernter Vergangenheit war Korruption noch gesellschaftlich akzeptiert. Diese Akzeptanz spiegelte sich bspw. in der Tatsache wider, dass Korruptionsgelder bis 1999 von der Steuer absetzbar waren (vgl. Altmeier, 2012). Große Korruptionsskandale, wie sie hier untersucht werden, haben zur Folge, dass die Konzerne aus wirtschaftlichem Interesse ebenso wie aus Reputationsmotivation transparenter und offensiver mit dem Thema umgehen.

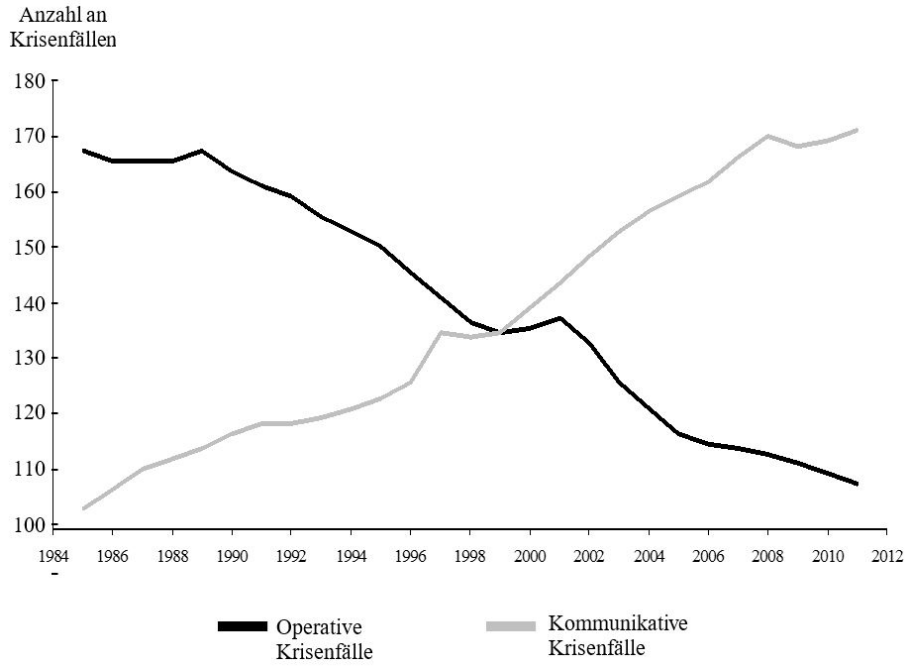

Abbildung 18: Entwicklung der Anzahl an Unternehmenskrisen (Deutschland, Schweiz, Österreich) - Steigende Fieberkurve (Quelle: In Anlehnung an Knauß, 2012; Institut für Krisenforschung, 2014)

Tabelle 3: Übersicht untersuchter potenzieller Krisenpräventionsmaßnahmen

(Eigene Darstellung)

Unternehmen	**Krisenjahr**	**Thema**	**Ratio** (Anzahl Hits je Krise/Dekaden-Median)
MAN1 (K2)	2009	Korruption	44311%
Deutsche Post4 (K1)	2013	Spionage	13360%
Siemens4 (K1)	2007	Korruption	13156%
VW1 (K1)	2005	Korruption + Lustreisen	10578%
Linde2 (K2)	2011	Korruption	6888%
SAP1 (K2)	2007	Spionage	6356%
Daimler1 (K2)	2004	Korruption	4391%
Deutsche Post2 (K1)	2004	Spionage + Wettbewerbsverzerrung	3236%
Siemens6 (K1)	2011	Bestechung	3096%
Deutsche Post3 (K1)	2011	Korruption	2336%
VW2 (K1)	2011	Korruption	1944%
Siemens2 (K1)	2004	Bestechung	1693%
E.ON1 (K2)	2006	Lustreisen	1627%
BASF3 (K2)	2010	Korruption	1256%
Deutsche Bank7 (K2)	2012	Dividenden-Stripping	1032%
Andere (K3)			< 1000%

Dass die Aufdeckungsrate solcher Fälle steigt, ist neben den genannten Entwicklungen auch auf die gesteigerte Sensibilität sowie das erhöhte Misstrauen der Menschen zurückzuführen (vgl. Altmeier, 2012). Der Corruption Perceptions Index (CPI) von Transparency International listet Länder nach dem Grad auf, in dem dort Korruption bei Amtsträgern, Politikern oder Unternehmen wahrgenommen wird. Laut Transparency International geben die Deutschen seit Jahren an, dass sie das Gefühl haben, die Korruption steige hierzulande (vgl. Transparency, 2014; Abbildung 20).

Deutschland befindet sich in dieser Listung auf Rang vier hinter den Niederlanden, der Schweiz und Belgien. Die potenzielle Neigung zur Bezahlung von Schmiergeldern reflektiert sich folglich in diesem Ranking.

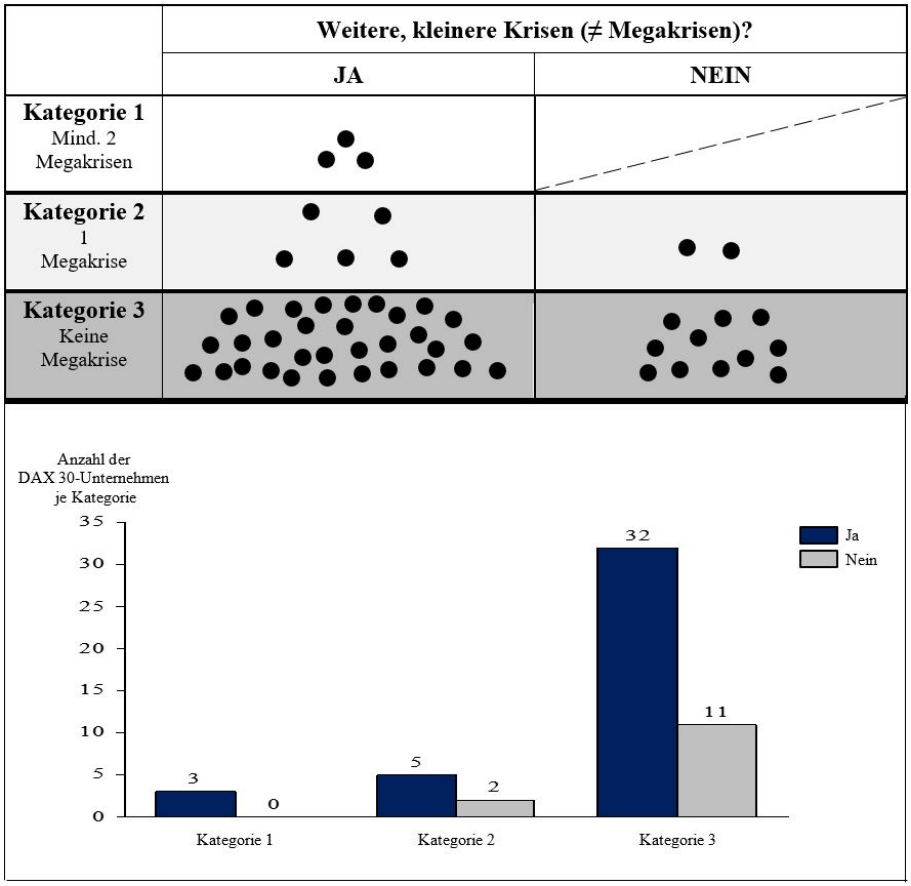

Abbildung 19: Kategorisierungsmatrix der DAX 30 Krisenfälle (Quelle: eigene Darstellung)

Die Organisation Transpareny International verpflichtet ihre (kooperativen) Mitglieder zur Einhaltung gewisser Regeln und Kodizes. Aktuelle (relevante) kooperative Mitglieder sind Allianz SE, BASF SE, Commerzbank AG, Daimler AG, Deutsche Lufthansa AG, MAN SE sowie SAP AG. Die Mitgliedschaft der Siemens AG wurde u.a. als Reaktion auf die Schmiergeldaffäre zum 15.12.2006 beendet.

3.2.3. DAX 30 krisenfreie Unternehmen

Wie bereits erwähnt, konnten für manche Unternehmen keine relevanten Krisenfälle identifiziert werden. Diese Unternehmen sind in der folgenden Tabelle 4 aufgeführt.

Tabelle 4 zeigt elf DAX 30-Unternehmen, welche als ethisch-moralisch krisenfrei identifiziert werden konnten. Bei eben diesen Unternehmen handelt es sich hauptsächlich um im B2B-Bereich tätige Unternehmen bzw. um Unternehmen, welche nicht im Fokus der Öffentlichkeit stehen. Anders als bei Kategorie 1-Unternehmen finden sich hier mit bspw. Epcos auch kleinere Unternehmen unter den DAX 30, welche grundsätzlich nicht im Übermaß mit Schlagzeilen auf sich aufmerksam gemacht haben.

Unternehmensumfirmierungen/ –Zusammenschlüsse wie bspw. die Bayerische Vereinsbank mit der Bayerischen Hypotheken- und Wechselbank im Jahr 1998 zur heutigen Hypovereinsbank zählen als neue Unternehmen. Im Zeitraum vor dem Zusammenschluss konnten am Beispiel der beiden Banken keine relevanten Unternehmenskrisen identifiziert werden.

3.2.4. Zeitliche Muster ethisch-moralisch bedingter Unternehmenskrisen

Wie bereits erwähnt, soll neben der Betrachtung der absoluten Anzahl der Krisen je Unternehmen auch die zeitliche Struktur der Krisen genauer betrachtet werden. Diese Betrachtungsweise ist im Besonderen aus der Perspektive der Pfadabhängigkeitstheorie interessant. Zu erwarten wäre, dass zeitlich aufsteigende Muster festzustellen sind, ausgehend von der Annahme, dass sich eine Unternehmenskrise durch diverse Anzeichen ankündigt. Findet kein Entgegenwirken des Unternehmens statt, d.h. es bleibt auf dem sich ausbildenden Pfad, so könnte dies in eine Megakrise führen. Jedoch lässt die Theorie auch Raum für andere Interpretationen. Die folgende Abbildung 21 21 zeigt die Chronologie bzw. den zeitlichen Verlauf identifizierter ethisch-moralisch bedingter Unternehmenskrisen auf:

Erwartet wurde, dass je Kategorie, oder zumindest für gewisse Gruppen von Unternehmen, ein Muster der zeitlichen Struktur zu erkennen ist. Gemäß der Pfadabhängigkeitstheorie wäre zu erwarten gewesen, dass sich eine Krise bspw. durch Warnsignale (erste kleine Krisen) abzeichnen würde. Solange die Diskrepanz zwischen Innen- und Außenperspektive noch nicht allzu groß ist, sollten maximal

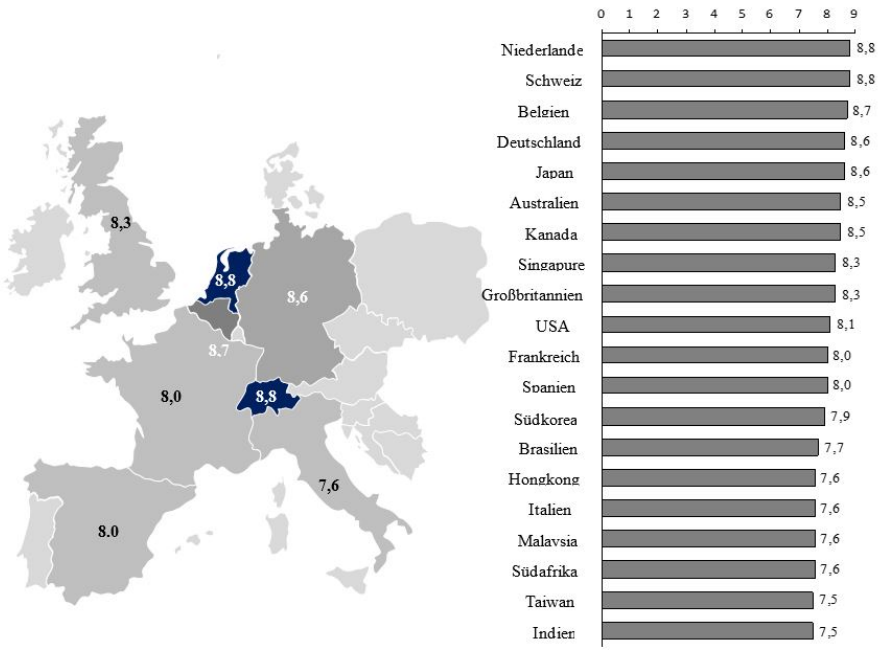

Abbildung 20: Bereitschaft zur Zahlung von Korruptionsgeldern 2011 (Quelle: In Anlehnung an Transparency International 2011)

Tabelle 4: Ethisch-moralisch „krisenfreieÜnternehmen in Deutschland im Zeitraum 1988 – 2014

(Eigene Darstellung)

Unternehmen	**Branche**
Altana	Chemiekonzern
Bayerische Vereinsbank	Financial Services (Bank)
Bayerische Hypotheken- und Wechselbank	Financial Services (Bank)
Deutsche Börse	Financial Services
Epcos	Hersteller elektronischer Bauelemente
Feldmühle Nobel	Papier-Chemie -Konzern
Hannover Rück	Financial Services (Versicherung)
K+S	Bergbauunternehmen
Salzgitter	Stahlkonzern
Schering	Pharmaunternehmen
VIAG Grundstoffe	Industrieunternehmen

kleine Krisen auftauchen. Schafft das Unternehmen es dann nicht, die Warnsignale wahrzunehmen, den Pfad zu verlassen, ist mit einer Megakrise zu rechen. Die Ergebnisse zeigen jedoch, dass es eben kein solches idealtypisches Krisenmuster zu geben scheint. Ethisch-moralisch bedingte Unternehmenskrisen können diesen Daten zufolge zu jeder Zeit, auch ohne vorherige Ankündigung durch bspw. kleinere Krisen, auftreten. Eine verallgemeinernde Vorhersage zeitlicher Krisenmustern und –Verläufen ist damit auf Grundlage dieser Daten für DAX 30-Unternehmen nicht möglich. Werden einzelne Unternehmen betrachtet, wie bspw. die Deutsche Bank, so könnten zuvor geäußerte Erwartungen in der Struktur interpretiert werden. Im Fall der Deutschen Bank ist ein aufbauender Verlauf der Krisen zu erkennen. Ein derartiges Muster könnte im Rahmen der Pfadabhängigkeitstheorie bedeuten, dass das Unternehmen kleinere Krisen erleidet, den Pfad jedoch beibehält und es erst nach einer Megakrise schafft, den Pfad zu verlassen. Jedoch reichen die vorliegenden Daten in diesem Fall nicht aus, um reliable Zukunftsprognosen abzugeben. Für MAN und SAP könnte gemäß der Pfadabhängigkeitstheorie die Schlussfolgerung bis zum heutigen Datum gezogen werden, dass, wenn ein Pfad bestand, dieser erfolgreich verlassen wurde und die Unternehmen keinen bloßen Pfadwechsel vollzogen haben.

3.2.5. Pfadabhängigkeitsauswertung

Wie bereits im Methodenabschnitt dieser Arbeit angekündigt, wird im Folgenden von einer Betrachtung der Einzel-

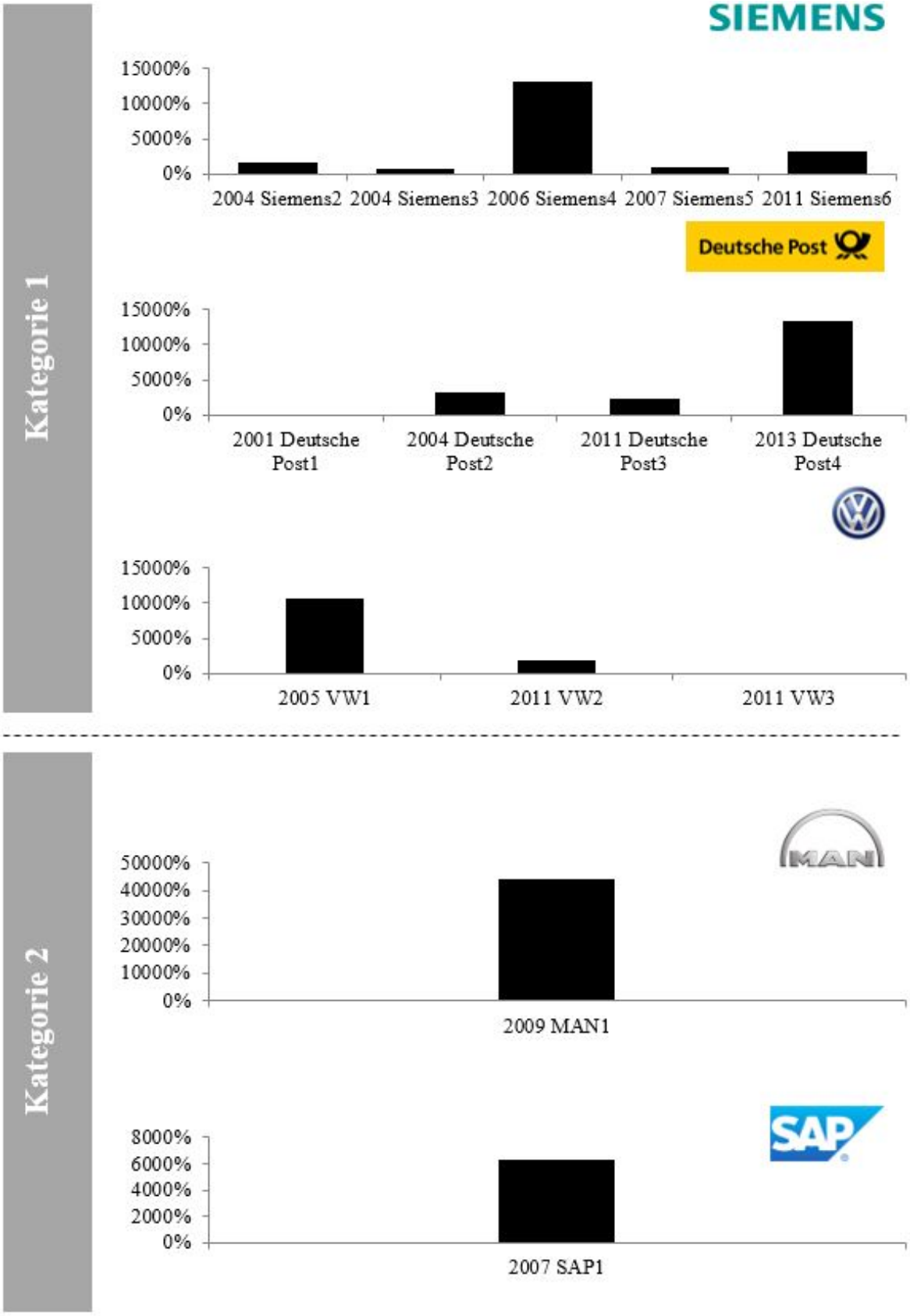

Abbildung 21: Chronologie ethisch-moralisch bedingter Unternehmenskrisen je Unternehmen (Deutschland) (Quelle: eigene Darstellung)

(Fortsetzung)

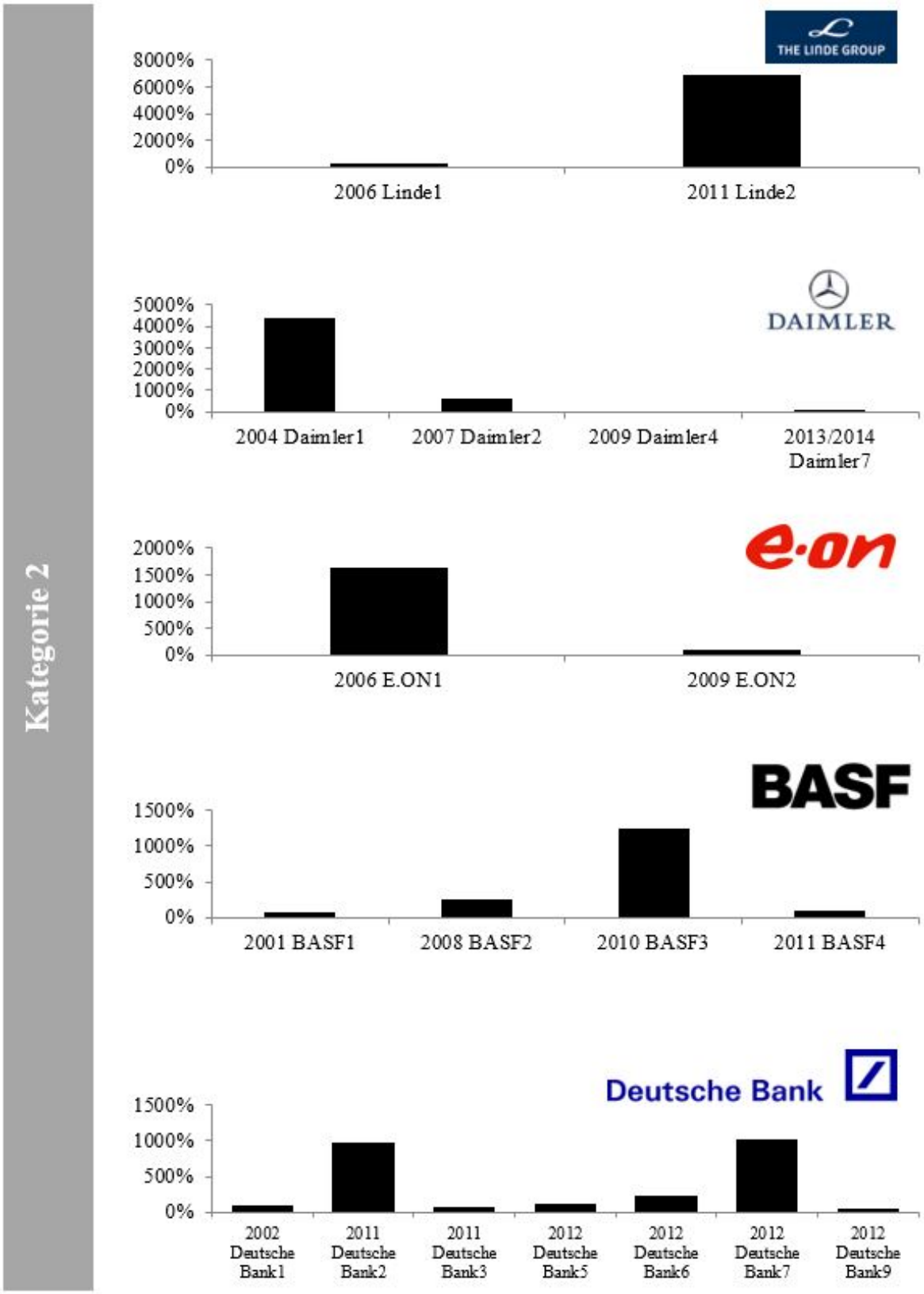

Abbildung 21—Fortsetzung

krisen auf eine Betrachtung auf Unternehmensebene übergegangen. Die Tabelle im Online-Appendix zeigt zum einen die Unterteilung aller identifizierten Unternehmen in krisenanfällige sowie –unanfällige Unternehmen. Zum anderen ist die binäre Kodierung (Ja = 1, Nein = 0) über das Zutreffen eines Kriteriums ersichtlich. Zum Verständnis der Rangbildung dient Abschnitt 3.1.3.

An dieser Stelle galt es herauszufinden, ob sich die krisenanfälligen von den krisenunanfälligen Unternehmen in den hier untersuchten Kriterien der Pfadabhängigkeit unterscheiden. Hierzu wird ein χ^2-Unabhängigkeitstest (2x2 Tabelle) durchgeführt mit den beiden Merkmalen „Krisenanfälligkeit von Unternehmen“ (Ausprägungen: anfällig und unanfällig) und den Kriterien der Pfadabhängigkeit (Ausprägungen: Persistenz, Ausmaß, Themenkonzentration). Die Tabellen 5, 6 und 7 zeigen den Berechnungsweg hin zu den jeweiligen χ^2-Werten – erwartete (statistisch berechnete) und beobachtete (tatsächlich gezählte) Werte, aus deren Abweichungen voneinander sich die jeweiligen χ^2-Werte errechnen lassen. Die ausführlichen Berechnungen finden sich in der beigefügten Excel-Tabelle.

Bei allen drei getesteten Kriterien der Pfadabhängigkeit lassen sich zum Teil große Abweichungen zwischen erwarteten und beobachteten Werten feststellen. Diese Diskrepanzen spiegeln sich auch in den Ergebnissen des durchgeführten χ^2-Unabhängigkeitstests wider, welche in Tabelle 8 aufgeführt sind.

Die These der Unabhängigkeit der Merkmale „Krisenanfälligkeit von Unternehmen“ und „Persistenz“ sowie „Ausmaß“ kann auf einem 0,1-Prozent Signifikanzniveau abgelehnt werden. Für das Merkmal „Themenkonzentration“ kann diese These auf einem 0,5-Prozent Signifikanzniveau verworfen werden. Dies deutet auf eine vorhandene Korrelation zwischen dem jeweiligen Kriterium der Pfadabhängigkeit und dem Merkmal der Krisenanfälligkeit hin. Um eine Aussage über die Stärke der soeben festgestellten Korrelation machen zu können, wurden ergänzend drei statistische Zusammenhangsmaße herangezogen, wobei die Ergebnisse des Cramer's V und des Phi-Koeffizienten aufgrund des Aufbaus als 2x2-Tabelle identisch sind. Sehr hohe Korrelationswerte kommen sehr selten vor. Krisenanfälligkeit von Unternehmen korreliert positiv mit den Kriterien der Pfadabhängigkeit. Damit kann die gemäß der Pfadabhängigkeitstheorie aufgestellte Vermutung, dass pfadabhängige Unternehmen krisenanfälliger sind als nicht pfadabhängige Unternehmen mithilfe dieser Ergebnisse für DAX 30-Unternehmen bekräftigt werden.

Ob es sich allerdings bei den untersuchten Krisenfällen um ein Pfadbrechen/ -verlassen oder einen Pfadwechsel handelt, sollte im Rahmen zukünftiger Forschung untersucht werden.

Die signifikante Korrelation der Merkmale „Krisenanfälligkeit von Unternehmen“ und den drei Kriterien der Pfadabhängigkeit eröffnet neue Ansätze zur Krisenprävention. So könnten Maßnahmen zur Vermeidung des Ausbildens einer Organisationalen Pfadabhängigkeit einen Ausgangspunkt für die Prävention ethisch-moralisch bedingter Unternehmenskrisen liefern. Weitere Ausführungen zu diesem Interpretationsansatz finden sich im Diskussionsteil dieser Arbeit.

3.2.6. Auswertung potenzieller Krisenpräventionsmaßnahmen

Wie auch schon im vorhergegangenen Abschnitt wird die Analyse effektiver Krisenpräventionsmaßnahmen auf Unternehmensebene durchgeführt. Die Tabelle im Online-Appendix zeigt die elf untersuchten potenziellen Krisenpräventionsmaßnahmen sowie die Einteilung aller DAX 30-Unternehmen in krisenanfällig und –unanfällig. Die Tabellen 9 bis 19 dokumentieren die elf χ^2-Unabhängigkeitstests je potenzieller Präventionsmaßnahme bzw. zeigen sowohl die in diesem Rahmen erwarteten und beobachteten Werte.

Auffällig an dieser Stelle zu beobachten ist, dass nur fünf aller Unternehmen überhaupt einen unabhängigen Krisenberater hinzugezogen haben und 52 dies nicht getan haben. Fälle wie die Siemens AG mit Theo Waigel oder auch die Daimler AG, welche sich als Reaktion auf die Korruptions-Affäre Christine Hohmann-Dennhardt, die ehemalige Richterin des Bundesverfassungsgerichts erst als Compliance-Beauftragte, dann in den Vorstand geholt haben, bleiben in Deutschland Einzelfälle.

Der deutsche Gesetzgeber ordnet nicht ausdrücklich an, dass Unternehmen ein Compliance-Management aufbauen müssen. Denn die Einhaltung des geltenden Rechts ist selbstverständlich. Neben branchenspezifischen Vorschriften (z.B. Umwelt- und Gesundheitsauflagen) gelten für jedes Unternehmen jedoch allgemeine Compliance-Anforderungen. Die generelle Compliance-Pflicht, Recht und Gesetz zu befolgen, schlägt sich z.B. in Standardeinhaltungs-/ Informations-/ oder Antragspflichten nieder (vgl. Bayrak et al., 2012: S. 8 f.).

In Bezug auf das Vorhandensein eines Compliance-Systems, zeigt sich, dass alle der krisenanfälligen und auch der Großteil nicht-krisenanfälliger Unternehmen diese Maßnahme durchführen. Dabei ist darauf hinzuweisen, dass sich unter den zehn Unternehmen ohne Compliance-System auch ältere Unternehmen befinden wie bspw. die Deutsche Babcock oder Hoechst, welche heute nicht mehr (in dieser Form) am Markt sind. Laut einer Studie von Pricewaterhouse Coopers verfügten im Jahr 2007 41 Prozent und im Jahr 2010 44 Prozent aller deutschen Großunternehmen über Compliance-Programme (vgl. Bussmann et al., 2010: S. 3). Wie die Wirtschaftsprüfungsgesellschaft KPMG in einem 2012 veröffentlichten Bericht aufzeigt, verfügte zu dieser Zeit bereits jedes zweite deutsche Großunternehmen über eine Compliance-Abteilung, welche als Ressort zumeist direkt den Vorständen unterstellt ist (vgl. Hartmann und Tauber, 2012).

Whistleblower Hotlines verbreiten sich zunehmend auch in Deutschland. Hierzulande ist deren Einrichtung zwar nicht wie in den USA für DAX 30-Unternehmen vorgeschrieben, jedoch treiben viele Unternehmen diese Entwicklung voran. Mit acht von zehn betreiben relativ gesehen mehr krisenanfällige Unternehmen ein Hinweisgebersystem, denn bei den unanfälligen Unternehmen sind es nur 26 von 47. Kategorienübergreifend verfügen rund 40 Prozent der Unternehmen

Tabelle 5: Persistenz

Beobachtete vs. Erwartete Werte beim χ^2-Unabhängigkeitstest der DAX 30-Unternehmen

			Persistenz		Gesamt
			Ja	Nein	
Krisenanfälligkeit von Unternehmen	Anfällig	Beobachter Wert	10	0	10
		Erwarteter Wert	4,21	5,79	10
	Unanfällig	Beobachter Wert	14	33	47
		Erwarteter Wert	19,79	27,21	47
Gesamt		Beobachtet Gesamt	24	33	57
		Erwartet Gesamt	24,0	33,0	57,0

Tabelle 6: Ausmaß

Beobachtete vs. Erwartete Werte beim χ^2-Unabhängigkeitstest der DAX 30-Unternehmen

			Ausmaß		Gesamt
			Ja	Nein	
Krisenanfälligkeit von Unternehmen	Anfällig	Beobachter Wert	10	0	10
		Erwarteter Wert	4,04	5,96	10
	Unanfällig	Beobachter Wert	13	34	47
		Erwarteter Wert	18,96	28,04	47
Gesamt		Beobachtet Gesamt	23	34	57
		Erwartet Gesamt	23,0	34,0	57,0

Tabelle 7: Themenkonzentration

Beobachtete vs. Erwartete Werte beim χ^2-Unabhängigkeitstest der DAX 30-Unternehmen

			Themenkonzentration		Gesamt
			Ja	Nein	
Krisenanfälligkeit von Unternehmen	Anfällig	Beobachter Wert	6	4	10
		Erwarteter Wert	2,28	7,72	10
	Unanfällig	Beobachter Wert	7	40	47
		Erwarteter Wert	10,72	36,28	47
Gesamt		Beobachtet Gesamt	13	44	57
		Erwartet Gesamt	13,0	44,0	57,0

über keine Whistleblower Hotline. Zu den meinst genannten Einwänden gegen die Einführung einer solchen Hotline zählt die Vermutung der möglichen Kommunikation von Misstrauen sowie die Erzeugung einer potenziellen Bespitzelungskultur wie die folgende Abbildung 23 zeigt.

Im Gegensatz zu Whistleblower Hotlines zeigt sich in Bezug auf Red-Flag Systeme, dass jeweils die Mehrheit der Unternehmen je Kategorie über kein solches System verfügt. Dies könnte darauf zurückgeführt werden, dass diese Art der Krisenprävention für manche Unternehmenstypen eher geeignet ist als für andere – diese Systeme finden sich häufiger in Industrieunternehmen wie Siemens als bspw. in einem Financial Services Unternehmen.

Ähnlich wie schon bei den Compliance-Systemen ist auch bzgl. Corporate Social Responsibility-Aktivitäten festzustellen, dass nahezu alle Unternehmen diese durchführen (49 von 57). Die Feststellung der reinen Aktivität sagt jedoch an dieser Stelle nichts über die Qualität, die Nachhaltigkeit oder die Glaubhaftigkeit dieser Maßnahmen aus.

In Bezug auf ein institutionalisiertes Rotationsprinzip ist festzuhalten, dass anteilig mehr nicht-krisenanfällige Unternehmen (14 von 33) Rotation betreiben im Vergleich zu den krisenanfälligen Unternehmen (zwei von acht). Insgesamt betrachtet scheint Rotation jedoch kein sehr weit verbreitetes Präventionskonzept zu sein, da nur 16 Unternehmen dies nutzen.

Tabelle 8: Ergebnisse des Unabhängigkeitstests – DAX 30 Pfadabhängigkeit

* Freiheitsgrade: (Anzahl Spalten – 1) * (Anzahl Zeilen – 1); Hier: 1
** zu einem 0,1%-Signifikanzniveau; χ^2 Wert: 10,83
*** zu einem 0,5%-Signifikanzniveau; χ^2 Wert: 7,88
**** Richtungstest des Phi-Koeffizienten in allen drei Fällen positiv → positive Korrelation der Variablen

	χ^2 Wert*	p-Wert	dF*	Entscheidung bzgl. Hypothese der Unabhängigkeit	Cramer's V	Kontingenz Koeffizient C	Phi-Koeffizient****
Persistenz	16,676**	0,00002	1	ablehnen	0,541	0,476	0,541
Ausmaß	17,928**	0,00001	1	ablehnen	0,561	0,489	0,561
Themenkonzentration	9,529***	0,00165	1	ablehnen	0,409	0,378	0,409

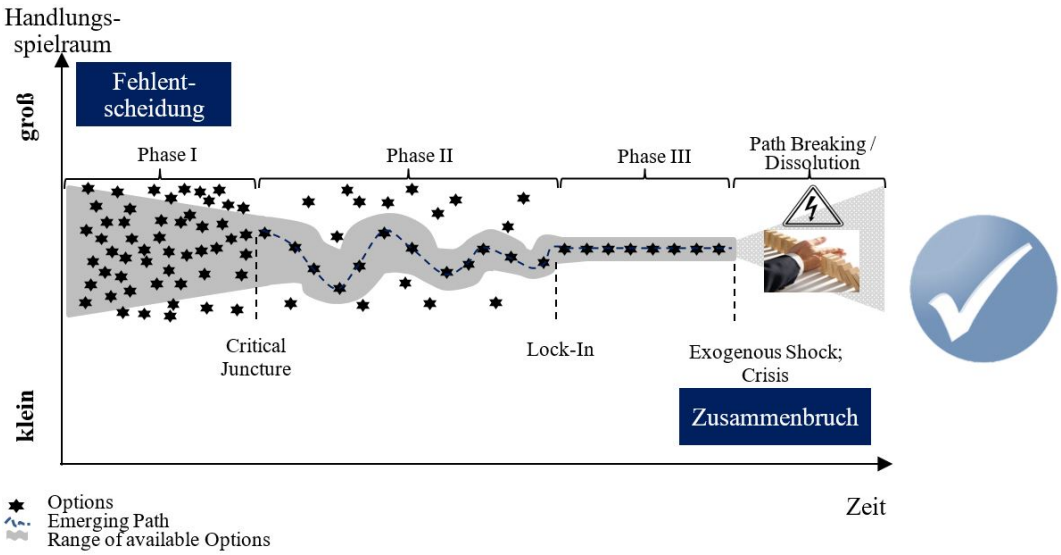

Abbildung 22: Ethisch-moralisch bedingte Unternehmenskrisen als Konsequenz Organisationaler Pfadabhängigkeit? Bei DAX 30-Unternehmen: Ja! (Quelle: eigene Darstellung)

Tabelle 9: Unabhängiger Krisenberater

Beobachtete vs. Erwartete Werte beim χ^2-Unabhängigkeitstest der DAX 30-Unternehmen

			Unabhängiger Krisenberater		Gesamt
			Ja	Nein	
Krisenanfälligkeit von Unternehmen	Anfällig	Beobachter Wert	2	8	10
		Erwarteter Wert	0,88	9,12	10
	Unanfällig	Beobachter Wert	3	44	47
		Erwarteter Wert	4,12	42,88	47
Gesamt		Beobachtet Gesamt	5	52	57
		Erwartet Gesamt	5,0	52,0	57,0

Auffällig ist hier, dass nahezu alle krisenanfälligen Unternehmen Reputationsmanagement betreiben. Bei den nicht-krisenanfälligen sind es nur die Hälfte der Unternehmen.

Rund die Hälfte aller Unternehmen betreiben Issues Management, wobei acht von zehn der krisenanfälligen und nur 19 der 47 krisenunanfälligen Unternehmen hier aktiv sind.

Für das Risikomanagement gelten die Erläuterungen wie bei Compliance-Systemen oder CSR-Aktivitäten.

Für das Corporate Governance Strukturen gelten die Erläuterungen wie bei Compliance-Systemen oder CSR-

Tabelle 10: Compliance-System

Beobachtete vs. Erwartete Werte beim χ^2-Unabhängigkeitstest der DAX 30-Unternehmen

			Compliance-System (Code of Conduct etc.)		Gesamt
			Ja	Nein	
Krisenanfälligkeit von Unternehmen	Anfällig	Beobachter Wert	10	0	10
		Erwarteter Wert	8,25	1,75	10
	Unanfällig	Beobachter Wert	37	10	47
		Erwarteter Wert	38,75	8,25	47
Gesamt		Beobachtet Gesamt	47	10	57
		Erwartet Gesamt	47,0	10,0	57,0

Tabelle 11: Whistleblower Hotline

Beobachtete vs. Erwartete Werte beim χ^2-Unabhängigkeitstest der DAX 30-Unternehmen

			Whistleblower Hotline		Gesamt
			Ja	Nein	
Krisenanfälligkeit von Unternehmen	Anfällig	Beobachter Wert	8	2	10
		Erwarteter Wert	5,96	4,04	10
	Unanfällig	Beobachter Wert	26	21	47
		Erwarteter Wert	28,04	18,96	47
Gesamt		Beobachtet Gesamt	34	23	57
		Erwartet Gesamt	34,0	23,0	57,0

Tabelle 12: Red-Flags System

Beobachtete vs. Erwartete Werte beim χ^2-Unabhängigkeitstest der DAX 30-Unternehmen

			Red-Flags System		Gesamt
			Ja	Nein	
Krisenanfälligkeit von Unternehmen	Anfällig	Beobachter Wert	4	6	10
		Erwarteter Wert	1,93	8,07	10
	Unanfällig	Beobachter Wert	7	40	47
		Erwarteter Wert	9,07	37,93	47
Gesamt		Beobachtet Gesamt	11	46	57
		Erwartet Gesamt	11,0	46,0	57,0

Aktivitäten.

Bei allen elf getesteten potenziellen Präventionsmaßnahmen lassen sich zum Teil große Abweichungen zwischen erwarteten und beobachteten Werten festestellen. Diese Diskrepanzen zeigen sich auch in den Ergebnissen des durchgeführten χ^2-Unabhängigkeitstests.

Entsprechend der Annahme einer krisenpräventiven Wirkung der elf untersuchten Maßnahmen wurde erwartet, dass die beiden Merkmale (je eine potenzielle Präventionsmaßnahme und die Krisenanfälligkeit von Unternehmen) signifikant negativ miteinander korrelieren. Entgegen dieser Erwartung ist jedoch festzustellen, dass nur vier der untersuchten potenziellen Krisenpräventionsmaßnahmen überhaupt mit dem Merkmal der Krisenanfälligkeit von Unternehmen signifikant korrelieren. Unternehmen, welche ein Red-Flags System betreiben oder Reputations-, Issues und Integrity Management durchführen, sind gemäß den deutschen Ergebnissen signifikant krisenanfälliger. Ob sie aber krisenanfälliger sind, weil sie die jeweiligen Maßnahmen betreiben oder diese Maßnahmen durchführen aufgrund ihrer Krisenanfälligkeit, kann nicht belegt werden. Über eine kausale Schlussfolgerung kann daher anhand dieser Ergebnisse keine Aussage getroffen werden. Die überraschender Weise positive Korrelation legt den Schluss nahe, dass deutsche Unternehmen

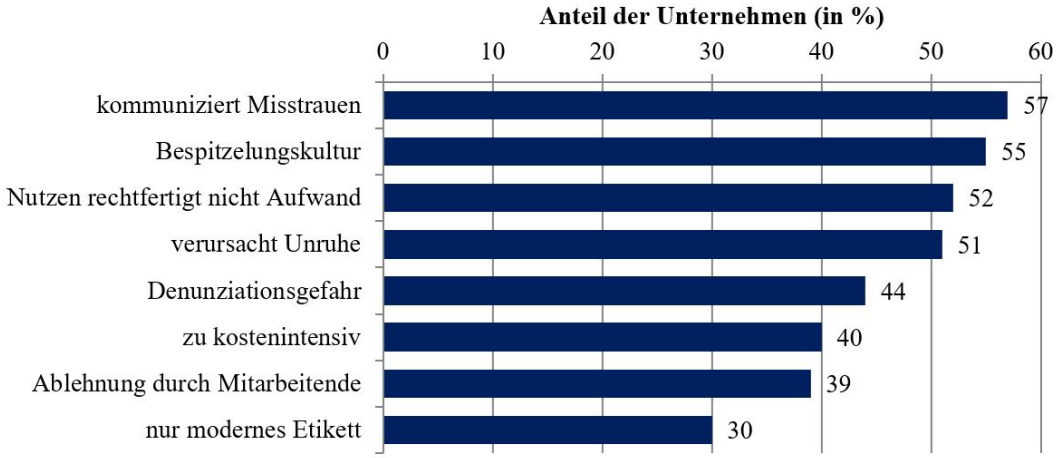

Abbildung 23: Einwände deutscher Großunternehmen gegen die Einführung einer Whistleblower Hotline (Quelle: Bussmann et al., 2010, S. 33)

Tabelle 13: Corporate Social Responsibility

Beobachtete vs. Erwartete Werte beim χ^2-Unabhängigkeitstest der DAX 30-Unternehmen

			CSR		Gesamt
			Ja	Nein	
Krisenanfälligkeit von Unternehmen	Anfällig	Beobachter Wert	10	0	10
		Erwarteter Wert	8,6	1,4	10
	Unanfällig	Beobachter Wert	39	8	47
		Erwarteter Wert	40,4	6,6	47
Gesamt		Beobachtet Gesamt	49	8	57
		Erwartet Gesamt	49,0	8,0	57,0

Tabelle 14: Rotationsprinzip

Beobachtete vs. Erwartete Werte beim χ^2-Unabhängigkeitstest der DAX 30-Unternehmen

			Rotationsprinzip		Gesamt
			Ja	Nein	
Krisenanfälligkeit von Unternehmen	Anfällig	Beobachter Wert	2	8	10
		Erwarteter Wert	2,81	7,19	10
	Unanfällig	Beobachter Wert	14	33	47
		Erwarteter Wert	13,19	33,81	47
Gesamt		Beobachtet Gesamt	16	41	57
		Erwartet Gesamt	16,0	41,0	57,0

keine strategische Krisenprävention betreiben. Sie setzen die Maßnahmen eher als Reaktion auf eine erlebte Krise ein.

Insgesamt sind die Ergebnisse zum 2,5/ 5 bzw. 10-Prozent Signifikanzniveau signifikant und die Stärke der Korrelation liegt bei maximal, 301. Dies kann u.a. durch die relativ geringe Stichprobengröße erklärt werden.

Unternehmen, welche intensiv z.B. Issues Management (Beispiel gewählt aufgrund des relativ höchsten Cramer's V-Werts) betreiben, haben tendenziell mehr Erfahrung mit Krisen und aufgrund dessen eine gewisse Aktivität in das Unternehmen integriert. Dies würde erklären, warum sich vor allem rein auf Krisenmanagement bezogene Abteilungen oder Aktivitäten vermehrt bei krisenanfälligen Unternehmen finden. Denn krisenunanfällige Unternehmen haben sich in der Vergangenheit nicht veranlasst gesehen, solche Maßnahmen zu ergreifen.

Betrachtet man über alle DAX 30-Unternehmen hinweg, welche Maßnahmen am populärsten sind, so trifft dies auf ein

Tabelle 15: Reputationsmanagement

Beobachtete vs. Erwartete Werte beim χ^2-Unabhängigkeitstest der DAX 30-Unternehmen

			Reputationsmanagement		Gesamt
			Ja	Nein	
Krisenanfälligkeit von Unternehmen	Anfällig	Beobachter Wert	9	1	10
		Erwarteter Wert	6,32	3,68	10
	Unanfällig	Beobachter Wert	27	20	47
		Erwarteter Wert	29,68	17,32	47
Gesamt		Beobachtet Gesamt	36	21	57
		Erwartet Gesamt	36,0	21,0	57,0

Tabelle 16: Issues Management

Beobachtete vs. Erwartete Werte beim χ^2-Unabhängigkeitstest der DAX 30-Unternehmen

			Issues Management		Gesamt
			Ja	Nein	
Krisenanfälligkeit von Unternehmen	Anfällig	Beobachter Wert	8	2	10
		Erwarteter Wert	4,74	5,26	10
	Unanfällig	Beobachter Wert	19	28	47
		Erwarteter Wert	22,26	24,74	47
Gesamt		Beobachtet Gesamt	27	30	57
		Erwartet Gesamt	27,0	30,0	57,0

Tabelle 17: Risikomanagement

Beobachtete vs. Erwartete Werte beim χ^2-Unabhängigkeitstest der DAX 30-Unternehmen

			Risikomanagement		Gesamt
			Ja	Nein	
Krisenanfälligkeit von Unternehmen	Anfällig	Beobachter Wert	9	1	10
		Erwarteter Wert	8,95	1,05	10
	Unanfällig	Beobachter Wert	42	5	47
		Erwarteter Wert	42,05	4,95	47
Gesamt		Beobachtet Gesamt	51	6	57
		Erwartet Gesamt	51,0	6,0	57,0

Compliance-System, CSR-Aktivitäten, Corporate Governance sowie Risiko-/Chancenmanagement zu. Diese Maßnahmen können als „Hygienefaktoren“ bezeichnet werden, da ihr Vorhandensein zwar „ein Muss“ zu sein scheint, jedoch zumindest in Bezug auf den Nutzen zu Krisenpräventionszwecken in Deutschland kein signifikanter Mehrwert festzustellen ist. Zu beleuchten wäre an dieser Stelle zum einen, welchen Einfluss die Qualität dieser Maßnahmen auf die Qualifizierung als Krisenpräventionsmaßnahme hat. Zum zweiten wäre es interessant zu untersuchen, welchen Einfluss deren Abwesenheit auf die Krisenanfälligkeit besitzt.

Auffällig ist des Weiteren, dass unter den DAX 30-Unternehmen Volkswagen, Daimler, Deutsche Bank, Deutsche Telekom, ThyssenKrupp, Bayer, Infineon, Deutsche Lufthansa, RWE und BMW nahezu alle untersuchten Maßnahmen (mind. neun von elf) durchführen. Damit zählen „die Großen der Großen“, d.h. die größten DAX-Konzerne, zu den aktivsten Maßnahmenimplementoren. Dies könnte zum einen damit zu erklären sein, dass diese aufgrund ihrer hohen Öffentlichkeitspräsenz einem starken Legitimationszwang ausgesetzt sind und sich daher aktiv bei der Implementierung (und Vermarktung) solcher Maßnahmen zeigen. Manche Unternehmen verfolgen mit der Einführung bestimmter Maßnahmen auch das Ziel eine Vorreiterrolle in Deutschland einzunehmen wie bspw. ThyssenKrupp mit der frühen Einführung einer Whistleblower Hotline im Jahr

Tabelle 18: Corporate Governance Strukturen

Beobachtete vs. Erwartete Werte beim χ^2-Unabhängigkeitstest der DAX 30-Unternehmen

			Corporate Governance Strukturen		Gesamt
			Ja	Nein	
Krisenanfälligkeit von Unternehmen	Anfällig	Beobachter Wert	10	0	10
		Erwarteter Wert	8,25	1,75	10
	Unanfällig	Beobachter Wert	37	10	47
		Erwarteter Wert	38,75	8,25	47
Gesamt		Beobachtet Gesamt	47	10	57
		Erwartet Gesamt	47,0	10,0	57,0

Tabelle 19: Integrity Management

Beobachtete vs. Erwartete Werte beim χ^2-Unabhängigkeitstest der DAX 30-Unternehmen

			Integrity Management		Gesamt
			Ja	Nein	
Krisenanfälligkeit von Unternehmen	Anfällig	Beobachter Wert	4	6	10
		Erwarteter Wert	1,75	8,25	10
	Unanfällig	Beobachter Wert	6	41	47
		Erwarteter Wert	8,25	38,75	47
Gesamt		Beobachtet Gesamt	10	47	57
		Erwartet Gesamt	10,0	47,0	57,0

2007 (vgl. Küttner, 2007). Darüber hinaus könnte eine Erklärung für die überproportionale Aktivität der großen DAX 30-Unternehmen sein, dass diese grundsätzlich mehr Kapital zur Verfügung haben, welches sie in entsprechende Maßnahmen investieren können. Zudem sind in manchen Fällen „Spillover-Effekte“ zu beobachten – wird bspw. die Einführung mancher Maßnahmen wie die Implementierung einer Whistleblower Hotline für Dow Jones 30-Unternehmen verpflichtend, so ist davon auszugehen, dass auch DAX 30-Konzerne, welche sowohl im deutschen als auch im amerikanischen Aktienindex gelistet sind wie z.B. Merck, diese Maßnahmen gesamtheitlich einführen. Die Ergebnisse zeigen jedoch klar, dass der Grundsatz „mehr ist besser“ bezüglich potenzieller Präventionsmaßnahmen nicht gilt. Die Menge durchgeführter Maßnahmen bildet folglich keinen besseren Schutz vor ethisch-moralisch bedingten Unternehmenskrisen. Vielmehr sollte hier in einem weiteren Schritt die Qualität implementierter Maßnahmen zusätzlich untersucht werden.

3.3. USA (Dow Jones 30)

In den folgenden sechs Abschnitten werden die erarbeiteten Ergebnisse für die Dow Jones 30-Unternehmen dargestellt. Eine Tabelle, welche die Dow Jones 30-Unternehmen seit 1988 darstellt, befindet sich im Online-Appendix. Eine Bewertung bzw. ein Vergleich mit den DAX 30-Ergebnissen findet im anschließenden Bewertungsabschnitt 3.4 statt.

3.3.1. Absolute Anzahl ethisch-moralisch bedingter Unternehmenskrisen

Wie auch schon für die Unternehmen innerhalb des DAX 30 Index, so soll nun auch für die Dow Jones 30-Unternehmen die Kategorisierung identifizierter Unternehmenskrisen in Abbildung 24 aufgezeigt werden.

- Es konnten drei Unternehmen in Kategorie 1, d.h. Unternehmen mit mindestens zwei Megakrisen zwischen 1988 und 2014, identifiziert werden. Diese drei Unternehmen sind namentlich Goldman Sachs, American Express und die Bank of America. Auffällig ist an dieser Stelle bereits, dass es sich bei den großen Krisenträgern um Banken bzw. Financial Services Anbieter handelt.
- Wie auch schon im DAX 30 kann aus der durchgeführten Auswertung auch für die Dow Jones 30-Unternehmen abgelesen werden, dass die Kategorie 1-Unternehmen allesamt neben den mindestens zwei Megakrisen noch weitere Krisen aufweisen. Diese Feststellung kann ein Indiz auf Pfadabhängigkeitsentwicklungen sein. Jedoch ist auch hier neben der Feststellung, dass es weitere, nicht den Megakrisen zugeordnete, Krisen gibt, auch die Chronologie bzw. das Muster der Krisen zu berücksichtigen. Diese systematische Betrachtung folgt.

Tabelle 20: Ergebnisse des Unabhängigkeitstests – DAX 30 Präventionsmaßnahmen

* Freiheitsgrade: (Anzahl Spalten – 1) * (Anzahl Zeilen – 1); Hier: 1
** zu einem 10%-Signifikanzniveau; kritischer χ^2 Wert: 2,71
*** zu einem 2,5%-Signifikanzniveau; kritischer χ^2 Wert: 5,02
**** zu einem 5%-Signifikanzniveau; kritischer χ^2 Wert: 3,84

	χ^2 **Wert**	**p-Wert**	**dF***	**Entscheidung bzgl. Hypothese der Unabhängigkeit**	**Cramer's V**	**Kontingenz Koeffizient C**	**Phi-Koeffizient**
Unabhängiger Krisenberater	1.911	0.159	1	Nicht ablehnen	0.183	0.18	0.183
Compliance-System/ Abteilung	2.58	0.106	1	Nicht ablehnen	0.213	0.208	0.213
Whistleblower Hotline	2.087	0.143	1	Nicht ablehnen	0.191	0.188	0.191
Red-Flags System	3.337**	0.068	1	ablehnen	0.242	0.235	0.242
CSR	1.98	0.153	1	Nicht ablehnen	0.186	0.183	0.186
Rotationsprinzip	0.391	0.395	1	Nicht ablehnen	0.083	0.083	0.083(-)
Reputationsmanagement	3.755**	0.053	1	ablehnen	0.257	0.249	0.257
Issues Management	5.180***	0.022	1	ablehnen	0.301	0.289	0.301
Risikomanagement	0.004	0.499	1	Nicht ablehnen	0.008	0.008	0.008
Corporate Governance Strukturen	2.58	0.106	1	Nicht ablehnen	0.213	0.208	0.213
Integrity Management	4.228****	0.04	1	ablehnen	0.272	0.263	0.272

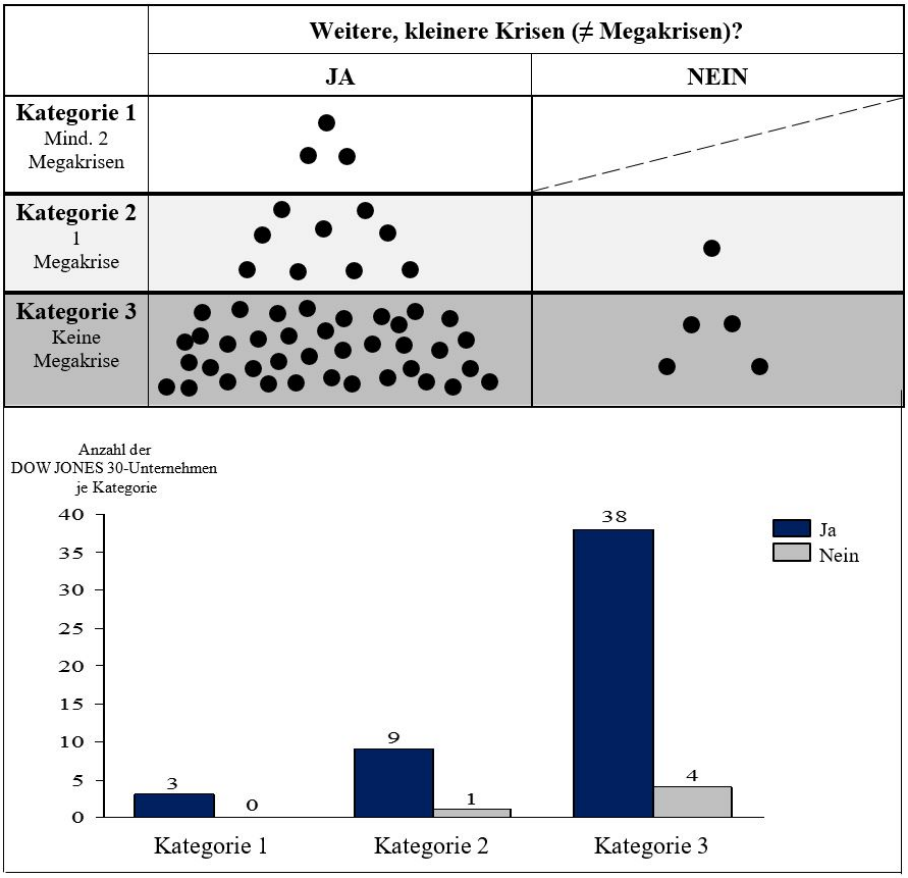

Abbildung 24: Kategorisierungsmatrix der Dow Jones 30 Krisenfälle (Quelle: eigene Darstellung)

- Kategorie 2 beinhaltet neun Unternehmen, welche neben der einen durchlebten Megakrise noch weitere, kleinere Krisen zu verzeichnen hatten. Im Gegensatz zum DAX 30 findet sich hier lediglich ein Unternehmen, welches zwar eine Megakrise, aber keine weiteren relevanten Krisen erlebt habt, namentlich Navistar. Hier wird die später folgende Betrachtung besonders interessant, da es herauszufinden gilt, welche Maßnahmen Navistar ergriffen hat, um damit weitere ethisch-moralisch bedingten Krisen vorzubeugen.
- In Kategorie 3 finden sich 38 Unternehmen, welche lediglich kleine ethisch-moralisch bedingte Krisen aufzuweisen haben. Interessant zum DAX 30 ist an dieser Stelle, dass die Gruppe von Unternehmen (Anzahl: vier), die weder einer Megakrise noch sonstige, kleinere ethisch-moralisch bedingte Krisen erlebt haben, im Vergleich zum DAX 30 (elf) sehr gering ist.

3.3.2. Dow Jones 30 „Megakrisen“

Die folgende Tabelle 21 zeigt das Ergebnis des soeben vorgestellten Identifikations- sowie Selektionsprozesses relevanter Krisenfälle.

Wie die Übersicht zeigt, finden sich überproportional häufig Banken-Krisenfälle unter den Dow Jones 30 Megakrisen. Dies lässt sich vor allem mit der Historie der Finanzmarkkrise, welche sich mitunter zuerst in den USA zeigte, erklären. Im Zuge des Post-Krisenmanagements verschärfen Aufsichtsbehörden die Richtlinien für Finanzinstitutionen weltweit zunehmend. Sogenannte pre-crisis mistakes (z.B. der Verkauf fauler Kredite oder die Manipulation von Richtzinssätzen wie dem Libor oder dem Euribor) stehen in den meisten Krisen- und Straffällen im Fokus (vgl. TV-Novosti, 2014) wie auch die steigenden verhängten Milliardenstrafen in Abbildung 25 zeigen.

Die Bank of America führt auch diese Rangliste der Strafzahlungen über Jahre hinweg an, was das Ergebnis der vorliegenden Analyse untermauert. Auch Finanzinstitutionen wie Goldman Sachs und Citigroup mussten in ihrer Vergangenheit oft wegen ethisch-moralisch bedingten Fehlverhaltes am Markt beträchtliche Strafen bezahlen. Neben der Höhe der jeweiligen Strafzahlungen hat, wie Abbildung 25 zeigt, auch die absolute Anzahl von Strafzahlungen seit der Finanzmarktkrise stark zugenommen.

3.3.3. Dow Jones 30 krisenfreie Unternehmen

Für manche Unternehmen innerhalb des Dow Jones 30 konnten zwischen 1988 und 2014 keine relevanten Krisenfälle identifiziert werden.

Tabelle 21: Die größten ethisch-moralisch bedingten Unternehmenskrisen in den USA im Zeitraum 1988 - 2014

K1 = Kategorie 1
K2 = Kategorie 2
K3 = Kategorie 3

Unternehmen	**Krisen-jahr**	**Thema**	**Ratio** (Anzahl Hits je Krise/ Dekaden-Median)
American Express4 (K1)	2012	Violation of federal law (marketing, billing)	3346%
Cisco5 (K2)	2013	Fraud (Patent Case)	3316%
GM1 (General Motors1) (K2)	2013	Fraud	2956%
Goldman8 (K1)	2013	Exchange Rate Manipulation	2806%
Goldman1 (K1)	2001	Illegal SWAP-Businesses	2049%
American Express1 (K1)	1994	Money Laundering	1959%
AT&T3 (K2)	2013	Violation of Privacy Protection	1890%
Verizon1 (K2)	2001	Telephone Slamming	1873%
Boeing1 (K2)	1989	Illegal actions (Pentagon planning documents)	1767%
Johnson2 (K2)	>1998	Bribery	1482%
BankOfAmerica3 (K1)	2011	Soured mortgage-backed securities	1448%
IBM1 (K2)	2003	Corruption	1325%
Goldman6 (K1)	2006	Aluminium, price manipulation	1296%
Cola1 (K2)	1999	Contaminated Coca-Cola beverages	1150%
Navistar1 (K2)	2007	Fraud	1139%
BankOfAmerica4 (K1)	2012	Libor Manipulation	1061%
BankOfAmerica1 (K1)	2008	Fraud (mortgages), Fannie Mae	1050%
Citigroup3 (K2)	2009	Fraud (mortgages)	1020%
Andere (K3)			< 1000%

Tabelle 22: Ethisch-moralisch ”krisenfreie” Unternehmen in den USA im Zeitraum 1988 – 2014

Unternehmen	**Branche**
Minnesota Mining & Manufacturing	Multi-Technologiekonzern
American Telephone & Telegraph Company	Telekommunikationsunternehmen
Allied-Signal	Luftfahrt-, Automobil- und Maschinenbauunternehmen
Union Carbide	Chemieunternehmen

Wie der Tabelle 22 zu entnehmen ist, konnten innerhalb des Dow Jones 30 nur vier ethisch-moralisch krisenfreie Unternehmen identifiziert werden. Bei eben diesen Unternehmen handelt es sich hauptsächlich um im B2B-Bereich tätige Unternehmen. In den Fällen Minnesota Mining & Manufacturing American und Telephone & Telegraph Company handelt es sich um Dow Jones 30-Unternehmen, welche zu einem späteren Zeitpunkt umfirmiert wurden in „3M“ bzw. „AT&T“. Wie im Methodik-Abschnitt erläutert, werden umfirmierte Unternehmen als neue Unternehmen behandelt. In dem Zeitpunkt, in dem die beiden hier angeführten Unternehmen also noch Mining & Manufacturing American und Telephone & Telegraph Company hießen und als solche im Dow Jones 30 Index gelistet waren, wurden keine ethisch-moralisch bedingten Krisen gefunden.

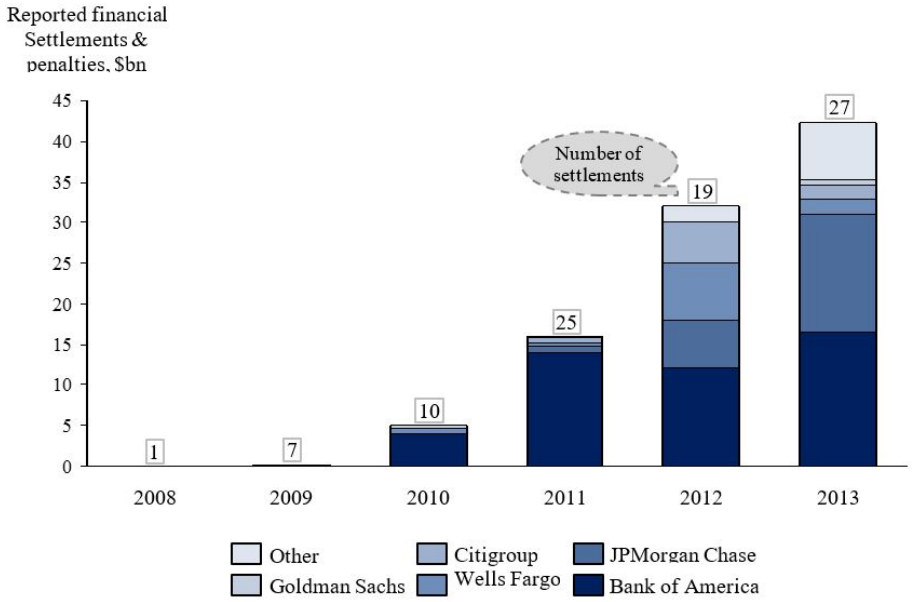

Abbildung 25: Reported financial settlements and penalties (Quelle: Economist, 2013)

3.3.4. Zeitliche Muster ethisch-moralisch bedingter Unternehmenskrisen

Auch für die Dow Jones 30 ethisch-moralisch bedingten Unternehmenskrisen soll eine Betrachtung der zeitlichen Struktur bzw. des Verlaufs der Krisen je Unternehmen vorgenommen werden. Eben diese Chronologie der Unternehmenskrisen zeigen die folgenden Abbildungen 26:

Erwartet wurde, dass je Kategorie oder zumindest für gewisse Unternehmen ein Muster der zeitlichen Struktur zu erkennen ist. Die Ergebnisse zeigen jedoch, dass es, wie bereits bei der Analyse der DAX 30 Daten, eben kein solche „Krisenmuster" zu geben scheint. Gemäß der Pfadabhängigkeitstheorie wäre zu erwarten gewesen, dass sich eine Krise durch Warnsignale (erste kleine Krisen) abzeichnen würde. Solange die Diskrepanz zwischen Innen- und Außenperspektive noch nicht allzu groß ist, sollten maximal kleine Krisen auftauchen. Schafft das Unternehmen es dann nicht, die Warnsignale wahrzunehmen und damit den Pfad zu verlassen, ist mit einer Megakrise zu rechen. Ein solcher idealtypischer Verlauf kann zumindest auf Kategorienebene nicht festgestellt werden. Ethisch-moralisch bedingte Unternehmenskrisen können diesen Daten zufolge zu jeder Zeit, auch ohne vorherige Ankündigung durch bspw. kleinere Krisen, auftreten. Eine verallgemeinernde Vorhersage von zeitlichen Krisenmustern und -Verläufen ist damit auf Grundlage dieser Daten auch für Dow Jones 30-Unternehmen nicht möglich.

3.3.5. Pfadabhängigkeitsauswertung

Die Kodier-Übersicht im Online-Appendix zeigt, wie bereits bei den DAX Unternehmen, die Einteilung der Dow Jones 30-Unternehmen in krisenanfällige und krisenunanfällige. Zudem wird das Kodier-Ergebnis der Kriterien der Pfadabhängigkeit (Persistenz, Ausmaß, Themenkonzentration) dargestellt.

An dieser Stelle galt es auch für Dow Jones 30-Unternehmen herauszufinden, ob sich die krisenanfälligen von den krisenunanfälligen Unternehmen in den hier untersuchten Kriterien der Pfadabhängigkeit unterscheiden. Hierzu wurde ein χ^2-Unabhängigkeitstest (2x2 Tabelle) durchgeführt mit den beiden Merkmalen „Krisenanfälligkeit von Unternehmen" (Ausprägungen: anfällig und unanfällig) und den Kriterien der Pfadabhängigkeit (Ausprägungen: Persistenz, Ausmaß, Themenkonzentration). Die Ergebnisse der einzelnen Kriterien der Pfadabhängigkeit sind den folgenden Tabellen zu entnehmen. Die Tabellen 23 bis 25 zeigen den „Berechnungsweg" hin zu den jeweiligen χ^2-Werten (erwartete und beobachtete Werte), aus deren Abweichungen voneinander sich die jeweiligen χ^2-Werten errechnen lassen. Die ausführlichen Berechnungen sind der beigefügten Excel-Tabelle zu entnehmen.

Alle drei getesteten Kriterien der Pfadabhängigkeit zeigen teils eine große Diskrepanz zwischen den erwarteten und den tatsächlich beobachteten Werten. Diese Abweichungen spiegeln sich auch in den Ergebnissen des durchgeführten χ^2-Unabhängigkeitstests wider, wie Tabelle 26 zu entnehmen ist.

Die Unabhängigkeitshypothese der Merkmale „Krisenanfälligkeit von Unternehmen" und allen drei Kriterien der Pfadabhängigkeit kann auf einem 0,1-Prozent Signifikanzniveau abgelehnt werden. Dieses Ergebnis legt eine Korrelation zwischen den Kriterien der Pfadabhängigkeit und dem Merkmal der Krisenanfälligkeit nahe. Für eine genauere Quantifizierung der Stärke der konstatierten positiven Korrelation wurden ergänzend drei statistisches Zusammenhangsmaße herangezogen, wobei die Ergebnisse des Cramer's V und des Phi-Koeffizienten aufgrund des Aufbaus als 2x2-Tabelle identisch sind. In den hier untersuchten Fällen handelt es sich um eine eher stark ausgeprägte positive Korrelation, da alle Cramer's

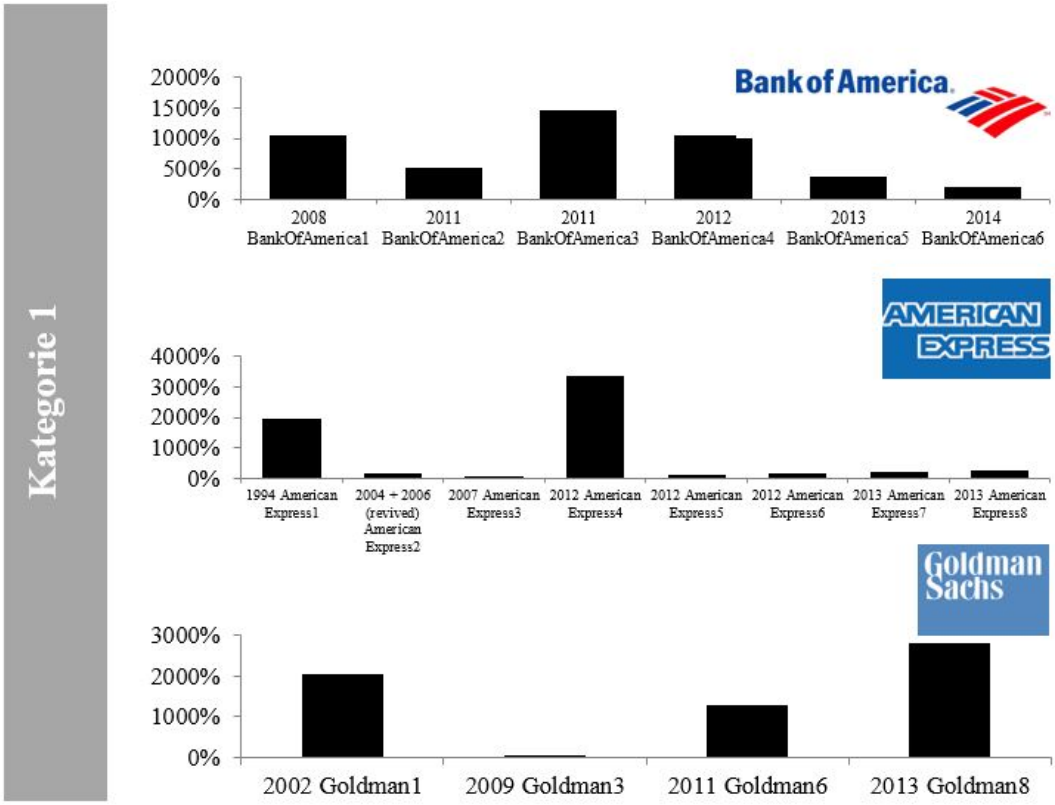

Abbildung 26: Chronologie ethisch-moralisch bedingter Unternehmenskrisen je Unternehmen (USA) (Quelle: eigene Darstellung)

(Fortsetzung)

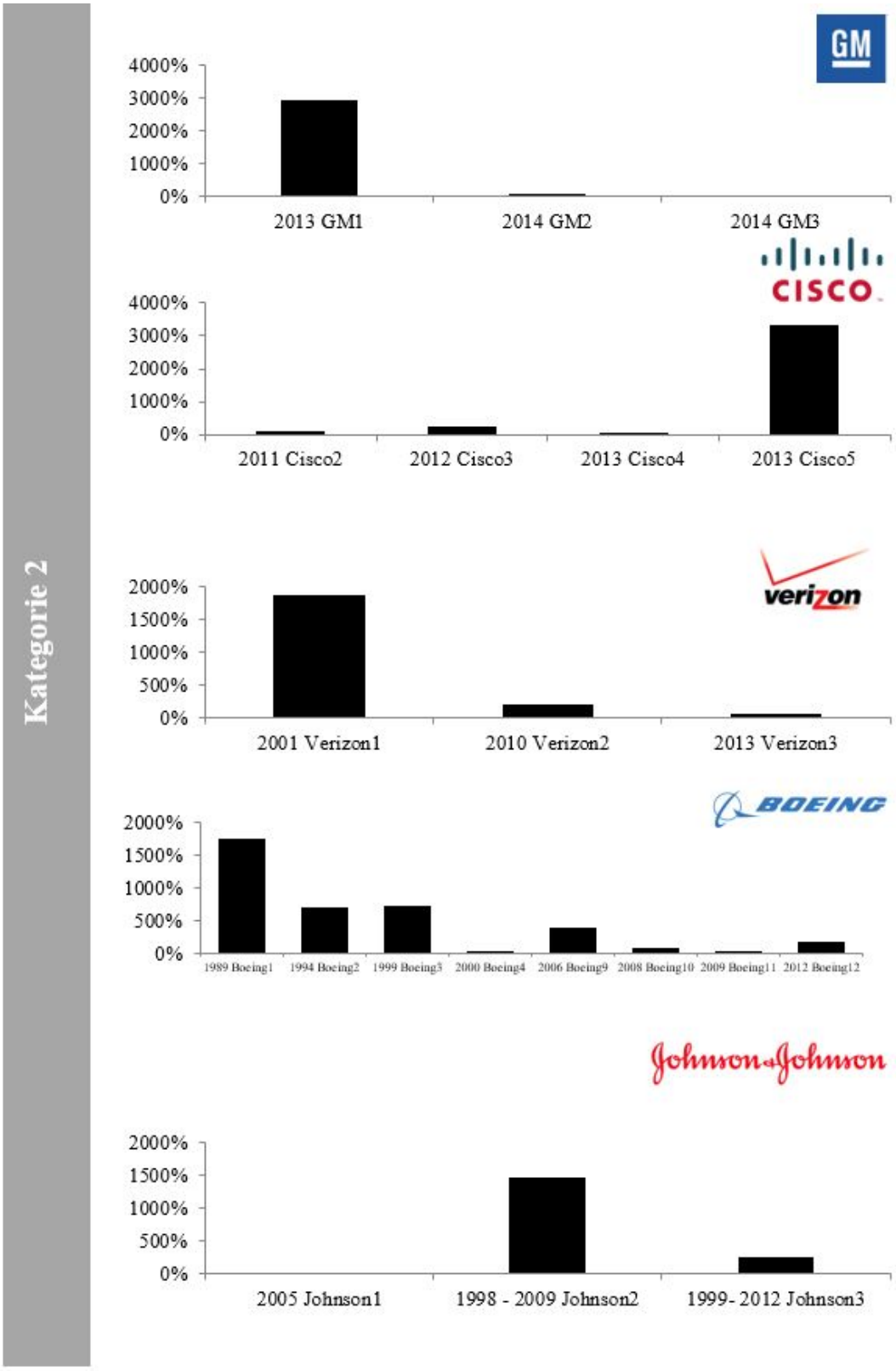

Abbildung 26—Fortsetzung

(Fortsetzung)

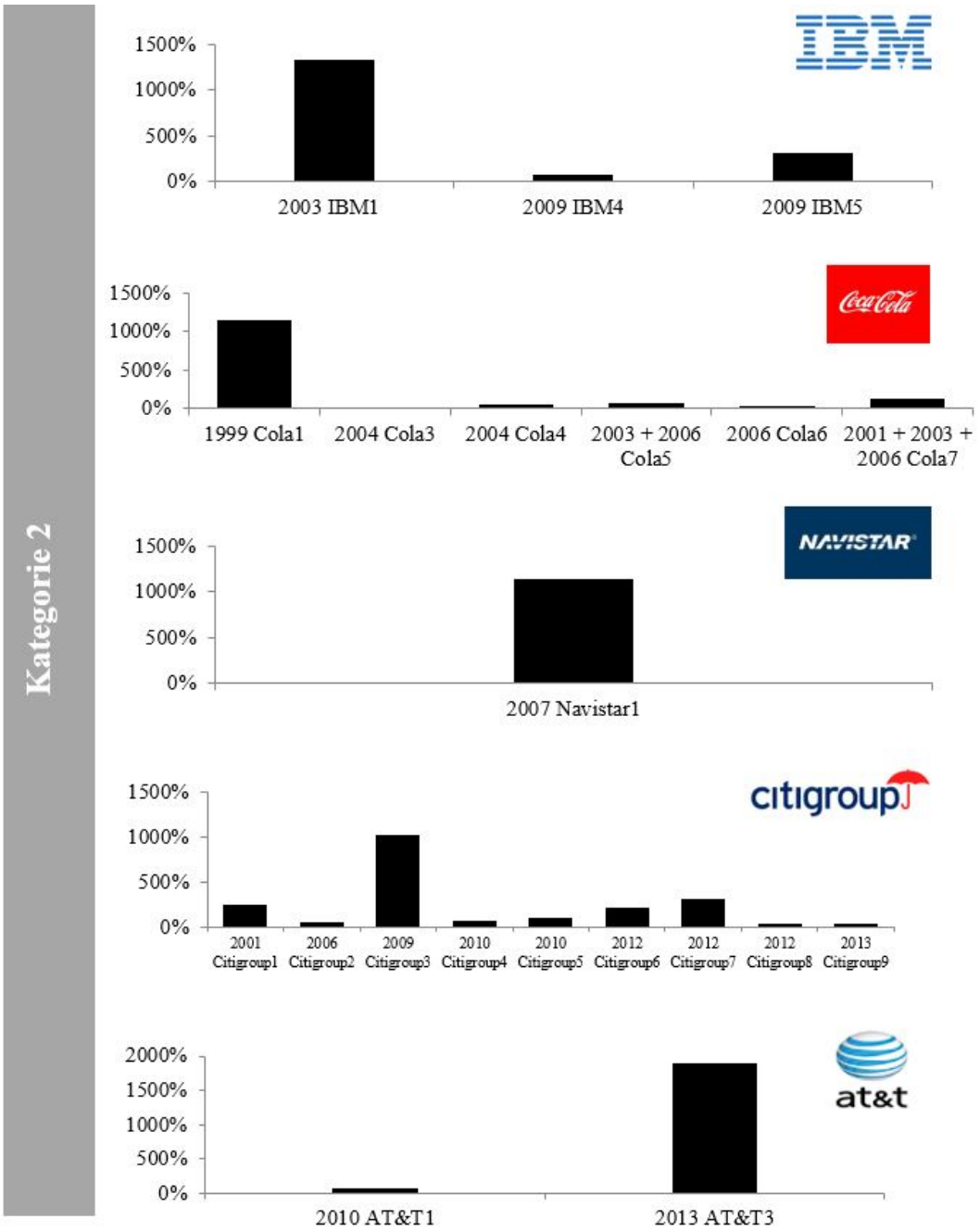

Abbildung 26—Fortsetzung

V-Werte größer als 0,5 sind. Krisenanfälligkeit von Unternehmen korreliert auch bei Dow Jones 30-Unternehmen positiv mit den Kriterien der Pfadabhängigkeit. So kann auch hier die gemäß der Pfadabhängigkeitstheorie aufgestellte Vermutung, dass pfadabhängige Unternehmen signifikant krisenanfälliger sind als nicht pfadabhängige Unternehmen, mithilfe dieser Ergebnisse bestätigt werden (siehe auch Abbildung 27).

Diese Erkenntnis liefert auch im Hinblick auf eine effektive Krisenprävention eine wichtige Basis. So könnten, wie auch schon für deutsche Unternehmen erwähnt, Maßnahmen zur Vermeidung pfadabhängiger Strukturen einen Ausgangspunkt für die Prävention ethisch-moralisch bedingter Unternehmenskrisen liefern. Weitere Ausführungen zu diesem Interpretationsansatz finden sich im Diskussionsteil dieser Arbeit.

3.3.6. Auswertung potenzieller Krisenpräventionsmaßnahmen

Wie auch schon im vorhergegangenen Abschnitt wird die Analyse effektiver Krisenpräventionsmaßnahmen auf Unternehmensebene durchgeführt. Die Tabelle im Online-Appendix zeigt die elf untersuchten potenziellen Krisen-

Tabelle 23: Persistenz

Beobachtete vs. Erwartete Werte beim χ^2-Unabhängigkeitstest der Dow Jones 30-Unternehmen

			Persistenz		Gesamt
			Ja	Nein	
Krisenanfälligkeit von Unternehmen	Anfällig	Beobachter Wert	20	0	20
		Erwarteter Wert	12,2	7,8	20
	Unanfällig	Beobachter Wert	16	23	39
		Erwarteter Wert	23,8	15,2	39
Gesamt		Beobachtet Gesamt	36	23	59
		Erwartet Gesamt	36,0	23,0	59,0

Tabelle 24: Ausmaß

Beobachtete vs. Erwartete Werte beim χ^2-Unabhängigkeitstest der Dow Jones 30-Unternehmen

			Ausmaß		Gesamt
			Ja	Nein	
Krisenanfälligkeit von Unternehmen	Anfällig	Beobachter Wert	20	0	20
		Erwarteter Wert	11,86	8,14	20
	Unanfällig	Beobachter Wert	15	24	39
		Erwarteter Wert	23,14	15,86	39
Gesamt		Beobachtet Gesamt	35	24	59
		Erwartet Gesamt	35,0	24,0	59,0

Tabelle 25: Themenkonzentration

Beobachtete vs. Erwartete Werte beim χ^2-Unabhängigkeitstest der Dow Jones 30-Unternehmen

			Themenkonzentration		Gesamt
			Ja	Nein	
Krisenanfälligkeit von Unternehmen	Anfällig	Beobachter Wert	13	7	20
		Erwarteter Wert	6,1	13,9	20
	Unanfällig	Beobachter Wert	5	34	39
		Erwarteter Wert	11,9	27,1	39
Gesamt		Beobachtet Gesamt	18	41	59
		Erwartet Gesamt	18,0	41,0	59,0

präventionsmaßnahmen sowie die Einteilung aller Dow Jones 30-Unternehmen in krisenanfällig und –unanfällig. Die folgenden Tabellen dokumentieren die elf χ^2-Unabhängigkeitstests je potenzieller Präventionsmaßnahme bzw. zeigen sowohl die in diesem Rahmen erwarteten als auch die beobachteten Werte.

Seit dem Jahr 2009 fordert die Obama-Regierung, dass in mindestens einem Drittel aller Bestechungs- und Korruptionsfälle Unternehmen trotz Zusatzkosten dazu verpflichtet werden, einen unabhängigen Beobachter oder Berater hinzuzuziehen (vgl. Pelofsky, 2012). Von allen Krisenursachen abstrahierend zeigt sich anhand der Ergebnisse, dass diese Forderung innerhalb der krisenanfälligen Unternehmen erfüllt ist, da sechs der 20 krisenanfälligen Unternehmen einen solchen Berater hinzugezogen haben bzw. dies fortwährend tun. Blickt man jedoch auf die krisenunanfälligen Unternehmen, ist festzustellen, dass nur acht von 39 einen unabhängigen Berater hinzugezogen haben bzw. dies fortwährend tun.

52 von 59 untersuchten Dow Jones 30-Unternehmen verfügen über ein Compliance-System. Diese sehr hohe absolute Anzahl ist u.a. auf die strengen US-Vorschriften des Sarbanes-Oxley Acts (SOX) zurückzuführen. Das Vorhandensein eines Compliance-Programms kann somit als Standard börsennotierter US-amerikanischer Unternehmen eingestuft werden.

SOX legt an der US-Börse gelisteten Unternehmen nahe, ein internes Verfahren einzurichten, über welches Mitarbei-

Tabelle 26: Ergebnisse des Unabhängigkeitstests – Dow Jones 30 Pfadabhängigkeit

* Freiheitsgrade: (Anzahl Spalten – 1) * (Anzahl Zeilen – 1); Hier: 1
** zu einem 0,1%-Signifikanzniveau; kritischer χ^2-Wert: 10,83
*** Richtungstest des Phi-Koeffizienten in allen drei Fällen positiv → positive Korrelation der Variablen

	c2 Wert	p-Wert	dF*	Entscheidung bzgl. Hypothese der Unabhängigkeit	Cramer's V	Kontingenz Koeffizient C	Phi-Koeffizient***
Persistenz	19,330**	0	1	ablehnen	0,572	0,497	0,572
Ausmaß	20,747**	0	1	ablehnen	0,593	0,51	0,593
Themenkonzentration	16,978**	0,00002	1	ablehnen	0,536	0,473	0,536

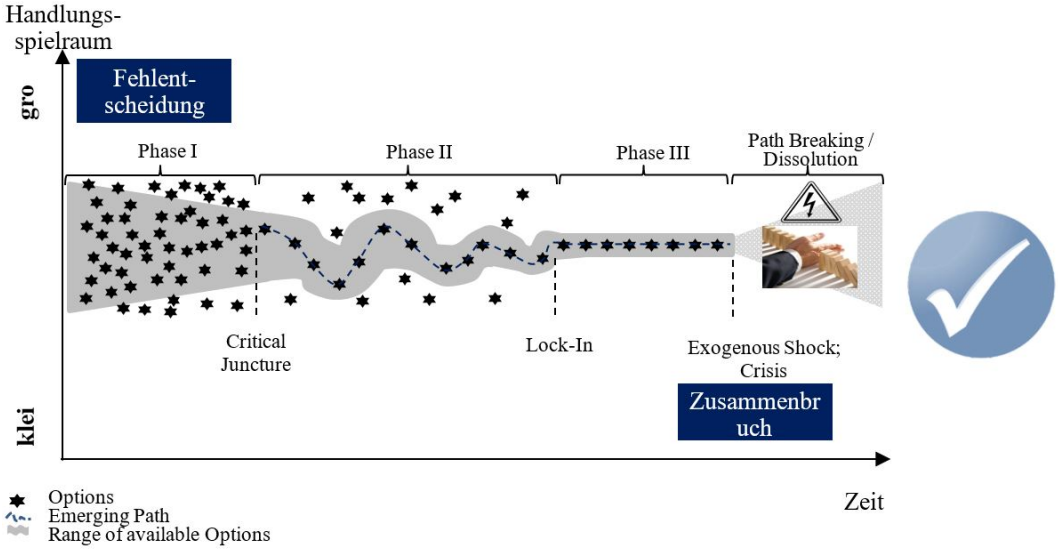

Abbildung 27: Ethisch-moralisch bedingte Unternehmenskrisen als Konsequenz Organisationaler Pfadabhängigkeit? Bei Dow Jones 30-Unternehmen: Ja! (Quelle: eigene Darstellung)

Tabelle 27: Unabhängiger Krisenberater

Beobachtete vs. Erwartete Werte beim χ^2-Unabhängigkeitstest der Dow Jones 30-Unternehmen

			Unabhängiger Krisenberater		Gesamt
			Ja	Nein	
Krisenanfälligkeit von Unternehmen	Anfällig	Beobachter Wert	6	14	20
		Erwarteter Wert	4,75	15,25	20
	Unanfällig	Beobachter Wert	8	31	39
		Erwarteter Wert	9,25	29,75	39
Gesamt		Beobachtet Gesamt	14	45	59
		Erwartet Gesamt	14,0	45,0	59,0

ter Fehlverhalten melden können. Immerhin 41 der insgesamt 59 Unternehmen verfügen über ein solches anonymes Hinweisgebersystem. Anzumerken ist, dass unter den krisenanfälligen Unternehmen 80 Prozent und nur 64 Prozent der krisenunanfälligen Unternehmen ein solches System implementiert haben.

Grundsätzlich ist anzumerken, dass die große Mehrheit der Dow Jones 30-Unternehmen nicht über ein Red-Flags System verfügt. Mögliche Erklärungen für die geringe Implementierungsrate unter US-Unternehmen finden sich im an-

Tabelle 28: Compliance-System / Abteilung

Beobachtete vs. Erwartete Werte beim χ^2-Unabhängigkeitstest der Dow Jones 30-Unternehmen

			Compliance-System / Abteilung		Gesamt
			Ja	Nein	
Krisenanfälligkeit von Unternehmen	Anfällig	Beobachter Wert	19	1	20
		Erwarteter Wert	17,63	2,37	20
	Unanfällig	Beobachter Wert	33	6	39
		Erwarteter Wert	34,37	4,63	39
Gesamt		Beobachtet Gesamt	52	7	59
		Erwartet Gesamt	52,0	7,0	59,0

Tabelle 29: Whistleblower Hotline

Beobachtete vs. Erwartete Werte beim χ^2-Unabhängigkeitstest der Dow Jones 30-Unternehmen

			Whistleblower Hotline		Gesamt
			Ja	Nein	
Krisenanfälligkeit von Unternehmen	Anfällig	Beobachter Wert	16	4	20
		Erwarteter Wert	13,9	6,1	20
	Unanfällig	Beobachter Wert	25	14	39
		Erwarteter Wert	27,1	11,9	39
Gesamt		Beobachtet Gesamt	41	18	59
		Erwartet Gesamt	41,0	18,0	59,0

Tabelle 30: Red-Flags System

Beobachtete vs. Erwartete Werte beim χ^2-Unabhängigkeitstest der Dow Jones 30-Unternehmen

			Red-Flags System		Gesamt
			Ja	Nein	
Krisenanfälligkeit von Unternehmen	Anfällig	Beobachter Wert	4	16	20
		Erwarteter Wert	3,05	16,95	20
	Unanfällig	Beobachter Wert	5	34	39
		Erwarteter Wert	5,95	33,05	39
Gesamt		Beobachtet Gesamt	9	50	59
		Erwartet Gesamt	9,00	50,00	59,00

schließenden Abschnitt 3.4. Chevron bspw. führt Analysen mithilfe eines solchen Systems durch. Bezeichnet wird dieses System allerdings als Fraud Prevention Analyse.

Corporate Social Responsibility wird von der großen Menge betrachteter Dow Jones 30-Unternehmen durchgeführt. Auffällig ist jedoch, dass 19 von 20 der krisenanfälligen Unternehmen CSR betreiben und damit eine starke Aktivität festzustellen ist. Bei den krisenunanfälligen Unternehmen finden sich immerhin zehn Unternehmen, welche keine CSR-Aktivitäten betreiben wie z.B. JP Morgan oder die The Travelers Companies.

Gezielte Aufgaben – und Positionsrotation betreiben nur eine Minderheit der Dow Jones 30-Unternehmen. Auffällig ist hier, dass nur drei von 20 krisenanfälligen und insgesamt acht von 39 krisenunanfälligen Unternehmen dieses Rotationsprinzip umsetzen.

Die Mehrheit aller Dow Jones 30-Unternehmen betreibt Reputationsmanagement, wobei unter den krisenanfälligen Unternehmen eine relativ stärkere Aktivität festzustellen ist. Sears Holdings Corporation oder die Bank of America finden sich in Folge ihrer schwachen bis nicht vorhandenen Reputationsmanagementaktivitäten seit Jahren auf den letzten Rängen des jährlichen Harris Poll Reputation Rankings (vgl. Merritt, 2014).

Mit 80 Prozent betreibt die große Mehrheit krisenanfälliger Unternehmen Issues Management. Bei den krisenunan-

Tabelle 31: Corporate Social Responsibility

Beobachtete vs. Erwartete Werte beim χ^2-Unabhängigkeitstest der Dow Jones 30-Unternehmen

			CSR		Gesamt
			Ja	Nein	
Krisenanfälligkeit von Unternehmen	Anfällig	Beobachter Wert	19	1	20
		Erwarteter Wert	16,27	3,73	20
	Unanfällig	Beobachter Wert	29	10	39
		Erwarteter Wert	31,73	7,27	39
Gesamt		Beobachtet Gesamt	**48**	**11**	**59**
		Erwartet Gesamt	48,00	11,00	59,00

Tabelle 32: Rotationsprinzip

Beobachtete vs. Erwartete Werte beim χ^2-Unabhängigkeitstest der Dow Jones 30-Unternehmen

			Rotationsprinzip		Gesamt
			Ja	Nein	
Krisenanfälligkeit von Unternehmen	Anfällig	Beobachter Wert	3	17	20
		Erwarteter Wert	3,73	16,27	20
	Unanfällig	Beobachter Wert	8	31	39
		Erwarteter Wert	7,27	31,73	39
Gesamt		Beobachtet Gesamt	11	48	59
		Erwartet Gesamt	11,00	48,00	59,00

Tabelle 33: Reputationsmanagement

Beobachtete vs. Erwartete Werte beim χ^2-Unabhängigkeitstest der Dow Jones 30-Unternehmen

			Reputationsmanagement		Gesamt
			Ja	Nein	
Krisenanfälligkeit von Unternehmen	Anfällig	Beobachter Wert	17	3	20
		Erwarteter Wert	13,22	6,78	20
	Unanfällig	Beobachter Wert	22	17	39
		Erwarteter Wert	25,78	13,22	39
Gesamt		Beobachtet Gesamt	39	20	59
		Erwartet Gesamt	39,00	20,00	59,00

Tabelle 34: Issues Management

Beobachtete vs. Erwartete Werte beim χ^2-Unabhängigkeitstest der Dow Jones 30-Unternehmen

			Issues Management		Gesamt
			Ja	Nein	
Krisenanfälligkeit von Unternehmen	Anfällig	Beobachter Wert	16	4	20
		Erwarteter Wert	12,2	7,8	20
	Unanfällig	Beobachter Wert	20	19	39
		Erwarteter Wert	23,8	15,2	39
Gesamt		Beobachtet Gesamt	36	23	59
		Erwartet Gesamt	36,00	23,00	59,00

fälligen Unternehmen sind es hingegen nur rund 51 Prozent, was auf eine stärkere Aktivität krisenanfälliger Unternehmen hindeutet.

Risiko-/ Chancenmanagement sowie auch Corporate Governance können wie auch schon für die DAX 30-Unternehmen als Standard bezeichnet werden – nahezu alle Unternehmen, egal ob krisenanfällig oder –unanfällig, führen diese Maßnahmen durch.

Integrity Management scheint eine nicht sehr verbreitete Maßnahme unter den Dow Jones 30-Unternehmen zu sein, denn 49 von insgesamt 59 Unternehmen führen diese Maßnahme nicht durch.

Auch für Dow Jones 30-Unternehmen wurde gemäß der krisenpräventiven Wirkungserwartung der untersuchten Merkmale unterstellt, dass die Merkmale signifikant negativ miteinander korrelieren, dass also z.B. CSR-Aktivitäten negativ mit der Krisenanfälligkeit zusammenhängen. Auch für amerikanische Unternehmen kann beobachtet werden, dass nur drei der untersuchten potenziellen Krisenpräventionsmaßnahmen überhaupt mit dem Merkmal der Krisenanfälligkeit von Unternehmen signifikant korrelieren. Unternehmen, welche ein Reputationsmanagement, Issues Management und CSR durchführen, sind den Ergebnissen zufolge signifikant krisenanfälliger. Kausalitätsbezogene Aussagen sind auch für Dow Jones 30-Unternehmen nicht möglich, da die χ^2-Prüfstatistik ausschließlich Rückschlüsse auf Korrelationen erlaubt.

Insgesamt sind die Ergebnisse zum 5 bzw. 10-Prozent Signifikanzniveau signifikant und die Stärke der Korrelation liegt bei maximal ,286. Dies kann u.a. durch die relativ geringe Stichprobengröße erklärt werden.

Betrachtet man über alle Dow Jones 30-Unternehmen hinweg, welche Maßnahmen am populärsten sind, so trifft dies auf ein Compliance-System, CSR-Aktivitäten, Corporate Governance sowie Risiko-/ Chancenmanagement und zusätzlich auf Whistleblower Hotlines zu. Diese Maßnahmen können als Hygienemaßnahmen bezeichnet werden, da ihr Vorhandensein zwar „ein Muss“ zu sein scheint, jedoch zumindest in Bezug auf den Nutzen zu Krisenpräventionszwecken in den USA kein signifikanter Mehrwert festzustellen ist. Zu beleuchten wäre an dieser Stelle zum einen, welchen Einfluss die Qualität dieser Maßnahmen auf die Qualifizierung als Krisenpräventionsmaßnahme hat. Zum zweiten wäre es interessant zu untersuchen, welchen Einfluss deren Abwesenheit auf die Krisenanfälligkeit besitzt.

Bemerkenswert ist des Weiteren, dass unter den Dow Jones 30-Unternehmen Chevron, Coca Cola, ExxonMobil, IBM, Microsoft und Wal-Mart nahezu alle untersuchten Maßnahmen (mind. neun von elf) durchführen. Auch hier sind die größeren Unternehmen des Index am aktivsten bei der Maßnahmendurchführung. Dies könnte zum einen damit zu erklären sein, dass diese aufgrund ihrer hohen Öffentlichkeitspräsenz einem starken Legitimationszwang ausgesetzt sind und sich daher aktiv bei der Implementierung (und Vermarktung) solcher Maßnahmen zeigen. Manche Unternehmen verfolgen mit der Einführung bestimmter Maßnahmen auch das Ziel eine Vorreiterrolle einzunehmen. Darüber hinaus könnte eine Erklärung für die überproportionale Aktivität der großen Dow Jones 30-Unternehmen sein, dass diese grundsätzlich mehr Kapital zur Verfügung haben, welches sie in entsprechende Maßnahmen investieren können. Die Ergebnisse zeigen jedoch klar, dass der Grundsatz „mehr ist besser“ bezüglich potenzieller Präventionsmaßnahmen nicht gilt. Die Menge durchgeführter Maßnahmen bildet folglich keinen besseren Schutz vor ethisch-moralisch bedingten Unternehmenskrisen. Vielmehr sollte hier in einem weiteren Schritt die Qualität implementierter Maßnahmen zusätzlich untersucht werden.

3.4. Vergleichende Betrachtung der Ergebnisse (Deutschland/ USA)

Dieser Abschnitt soll vor allem aufzeigen, welche Unterschiede oder Gemeinsamkeiten sich in den deutschen und den amerikanischen Ergebnissen finden lassen.

3.4.1. Anzahl identifizierter Krisenfälle

In den USA konnte eine erheblich höhere Anzahl relevanter Krisenfälle identifiziert werden: 125 DAX 30 vs. 218 Dow Jones 30 Fälle relevanter ethisch-moralisch bedingter Unternehmenskrisen. Ein möglicher Erklärungsansatz für diesen Unterschied findet sich in der Suchstrategie: DAX 30-Fällen wurden in deutscher Sprache bzw. über die Eingabe deutscher Schlagworte in diversen Datenbanken gesucht. Dow Jones 30 Fälle wurden zwar in denselben Verzeichnissen, jedoch mithilfe englischer Schlagworte gesucht. Es ist davon auszugehen, dass die Eingabe englischer Schlagworte aufgrund der weitaus höheren Verbreitung von Englisch als „Weltsprache“ sich in der Anzahl identifizierter Fälle niederschlägt. Die Unterschiede in den Dekaden-Medianen aus Deutschland und den USA unterstützen diese These: Die Dow Jones Dekaden-Mediane liegen konsequent über den DAX Dekaden-Medianen, woraus auf eine höhere (englischsprachige) Publikationsaktivität geschlossen werden kann.

3.4.2. Krisenthema

Verschiedene ethisch-moralische Hintergründe der Krisen konnten festgestellt werden – in Deutschland eher Korruptions-/ Spionagefälle, in den USA finden sich auffällig viele Bankenfälle (Betrug, Manipulation etc.). Dies kann sowohl auf unterschiedlich ablaufende gesellschaftliche Entwicklungen als auch zum Teil auf politische Rahmenbedingungen zurückgeführt werden. Der Foreign Corrupt Practices Act (FCPA) untersagt in den USA bereits seit 1977 jegliche Art der Korruption, wohingegen ein entsprechendes Gesetz in Deutschland erst 1999 erlassen wurde (vgl. Bayrak et al., 2012: S. 18). Dies liefert eine Erklärung in Bezug auf die in Deutschland gegenüber den USA sehr viel häufiger identifizierten Korruptionsfälle. Auch die in Deutschland relativ häufig auftauchenden durch Spionage bedingten Krisenfälle lassen sich durch „soziale Empfindlichkeiten“ erklären. Deutschland traf der NSA-Skandal 2013 härter, stieß auf mehr Überraschung und Empörung als in den USA. Dies kann u.a. darauf zurückgeführt werden, dass in den USA (vor allem seit

Tabelle 35: Risikomanagement

Beobachtete vs. Erwartete Werte beim χ^2-Unabhängigkeitstest der Dow Jones 30-Unternehmen

			Risikomanagement		Gesamt
			Ja	Nein	
Krisenanfälligkeit von Unternehmen	Anfällig	Beobachter Wert	19	1	20
		Erwarteter Wert	17,63	2,37	20
	Unanfällig	Beobachter Wert	33	6	39
		Erwarteter Wert	34,37	4,63	39
Gesamt		Beobachtet Gesamt	52	7	59
		Erwartet Gesamt	52,00	7,00	59,00

Tabelle 36: Corporate Governance Strukturen

Beobachtete vs. Erwartete Werte beim χ^2-Unabhängigkeitstest der Dow Jones 30-Unternehmen

			Corporate Governance Strukturen		Gesamt
			Ja	Nein	
Krisenanfälligkeit von Unternehmen	Anfällig	Beobachter Wert	19	1	20
		Erwarteter Wert	16,95	3,05	20
	Unanfällig	Beobachter Wert	31	8	39
		Erwarteter Wert	33,05	5,95	39
Gesamt		Beobachtet Gesamt	50	9	59
		Erwartet Gesamt	50,00	9,00	59,00

Tabelle 37: Integrity Management

Beobachtete vs. Erwartete Werte beim χ^2-Unabhängigkeitstest der Dow Jones 30-Unternehmen

			Integrity Management		Gesamt
			Ja	Nein	
Krisenanfälligkeit von Unternehmen	Anfällig	Beobachter Wert	4	16	20
		Erwarteter Wert	3,39	16,61	20
	Unanfällig	Beobachter Wert	6	33	39
		Erwarteter Wert	6,61	32,39	39
Gesamt		Beobachtet Gesamt	10	49	59
		Erwartet Gesamt	10,00	49,00	59,00

9/11) eine größere Akzeptanz von Überwachungs- und Spionageaktivitäten herrscht als in Deutschland. Obwohl in den USA, ebenso wie in Deutschland, Überwachung und Spionage seitens des Staates eine lange Historie haben, so ist die gefühlte gesellschaftliche Inakzeptanz in Deutschland weitaus größer, die Empörung und gesellschaftliche Unruhe bei aufgedeckten Spionagefällen höher als bei vergleichbaren Fällen in den USA.

3.4.3. Kategorien-Verteilung

Bei allen hier aufgeführten Unterschieden gleichen sich Kategorienschemen der Unternehmen bzgl. ihrer Verteilung in den drei Kategorien. In Kategorie 1 finden sich in beiden Märkten (USA und Deutschland) nur wenige Unternehmen. Unter diesen Kategorie 1-Unternehmen jedoch existiert keines, welches ausschließlich mindestens zwei Megakrisen, aber keine weiteren relevanten Krisen aufzuweisen hat. Es kann daraus geschlossen werden, dass nach dem Durchlaufen mehrerer Megakrisen die Wahrscheinlichkeit ethisch-moralisch krisenfrei zu werden gering ist. Kategorie 2 kann in beiden Märkten eine bereits wesentlich höhere absolute Anzahl an Unternehmen zugeordnet werden. Es zeigt sich zudem, dass sich hier auch Unternehmen finden lassen, welche zwar eine Megakrise durchlaufen haben, ansonsten jedoch ethisch-moralisch krisenfrei sind. Darauf beziehend kann abgeleitet werden, dass die Chance ethisch-moralisch

Tabelle 38: Ergebnisse des Unabhängigkeitstests – Dow Jones 30 Präventionsmaßnahmen

* Freiheitsgrade: (Anzahl Spalten – 1) * (Anzahl Zeilen – 1); Hier: 1
** zu einem 10%-Signifikanzniveau; kritischer χ^2-Wert: 2,71
*** zu einem 5%-Signifikanzniveau; kritischer Wert: 3,84

	c2 Wert	**p-Wert**	**dF***	**Entscheidung bzgl. Hypothese der Unabhängigkeit**	**Cramer's V**	**Kontingenz Koeffizient C**	**Phi-Koeffizient**
Unabhängiger Krisenberater	0,657	0,337	1	nicht ablehnen	0,106	0,105	0,106
Compliance-System / Abteilung	1,363	0,221	1	nicht ablehnen	0,152	0,15	0,152
Whistleblower Hotline	1,576	0,194	1	nicht ablehnen	0,163	0,161	0,163
Red-Flags System	0,527	0,364	1	nicht ablehnen	0,095	0,094	0,095
CSR	3,713**	0,054	1	ablehnen	0,251	0,243	0,251
Rotationsprinzip	0,265	0,427	1	nicht ablehnen	0,067	0,067	0,067(-)
Reputationsmanagement	4,823***	0,028	1	ablehnen	0,286	0,275	0,286
Issues Management	4,584***	0,032	1	ablehnen	0,279	0,268	0,279
Risikomanagement	1,363	0,221	1	nicht ablehnen	0,152	0,15	0,152
Corporate Governance Strukturen	2,461	0,114	1	nicht ablehnen	0,204	0,2	0,204
Integrity Management	0,2	0,443	1	nicht ablehnen	0,058	0,058	0,058

krisenfrei zu sein/werden steigt, umso weniger Megakrisen ein Unternehmen zu verzeichnen hat. In Kategorie 3 finden sich absolut gesehen in beiden Märkten ungefähr gleich viele Krisenfälle. In Deutschland sind jedoch elf krisenfreie Unternehmen, in den USA nur vier.

3.4.4. Anzahl krisenfreier Unternehmen

In Deutschland wurden elf, in den USA hingegen nur vier, Unternehmen als ethisch-moralisch krisenfrei eingestuft. Gemeinsam ist bei diesen Fällen zu konstatieren, dass es sich zumeist um im B2B-Bereich (z.B. Chemie, Pharma etc.) tätige Unternehmen mit geringer öffentlicher Präsenz handelt. Sind die Produkte oder Dienstleistungen der Unternehmen in erster Linie nicht für Endkunden relevant, so ist eine mögliche Hypothese, dass auch diese Unternehmen sich nicht jederzeit ethisch-moralisch korrekt verhalten. Jedoch aufgrund ihrer vergleichsweise geringeren Bedeutung für Endkunden bzw. das Gros der Gesellschaft bleiben diese Fälle eher unaufgedeckt oder generieren kaum bis keine Berichterstattung. Darüber hinaus bleibt die Auffälligkeit der geringeren Anzahl krisenfreier amerikanischer Unternehmen.

3.4.5. Verfolgung ethisch-moralischer Vergehen

Hier kann als Gemeinsamkeit festgestellt werden, dass in beiden Märkten ein hohes Aufklärungsinteresse vorherrscht und entsprechende Aufdeckungsaktivitäten stattfinden. Die Transparenzbestrebungen und die Nichttoleranz ethisch-moralischer Vergehen haben sowohl in den USA als auch in Deutschland erheblich zugenommen. Dies zeigt sich u.a. durch die überproportionale Präsenz aktueller Krisenfälle (Fälle nach dem Jahr 2000). Obwohl die Entstehungshistorie eben dieser gefahndeten und aufgedeckten Krisenfälle oft lange zurückgeht, was ein Indiz auf Pfadabhängigkeit ist, kam es häufig erst in unserer heutigen Zeit und Gesellschaft mit all ihren Werten zur Aufdeckung. Hinzu kommt, dass im Zeitalter der Globalisierung die Aufdeckungswahrscheinlichkeit tendenziell ansteigt, da Unternehmen sich immer häufiger auch den Ansprüchen anderer Länder auf Transparenz, soziales Verhalten oder Fairness stellen müssen. Bei nicht ethisch-moralisch einwandfreiem Verhalten laufen Unternehmen folglich Gefahr, enttarnt zu werden. Daher nehmen zumindest die sichtbaren Bestrebungen der Unternehmen, sich ethisch-moralisch sauber zu verhalten, zu. Werte und damit ganze Gesellschaften haben sich geändert und die strafenden Blicke auf solche mittlerweile als

moralische Vergehen eingestuften Gefälligkeiten haben sich drastisch verschärft. Es gibt weltweit immer weniger Möglichkeiten bei ethisch-moralisches Fehlverhalten ungestraft zu bleiben (vgl. Hartmann und Tauber, 2012).

3.4.6. Chronologie ethisch-moralisch bedingter Unternehmenskrisen

Weder für deutsche noch für amerikanische Unternehmen konnte ein zeitliches Krisenmuster festgestellt werden. Die Betrachtung des Vorkommens ethisch-moralischer Megakrisen und kleinerer Krisen je Kategorie lässt keine pauschale Prognose des Verlaufs einer solchen Unternehmenskrise zu. Ethisch-moralisch bedingte Unternehmenskrisen können in jedem Unternehmen entstehen und kündigen sich nicht zwangsläufig durch greifbare Anzeichen an. Eine Detailanalyse je Unternehmen könnte eine sinnvolle Ergänzung zur hier durchgeführten Analyse auf Kategorienniveau sein.

3.4.7. Pfadabhängigkeit

Die allgemeinen Charakteristika der Pfadabhängigkeit wurden im Rahmen dieser Arbeit anhand von drei Kriterien operationalisiert und untersucht (siehe Abbildung 28:

Anhand der vorliegenden Daten konnte sowohl für Deutschland als auch für die USA aufgezeigt werden, dass alle drei untersuchten Kriterien der Pfadabhängigkeit signifikant mit dem Merkmal der Krisenanfälligkeit von Unternehmen korrelieren. Damit kann im Sinne der Pfadabhängigkeitstheorie bzw. der daraus abgeleiteten Hypothese bestätigt werden, dass pfadabhängige Unternehmen eher eine ethisch-moralisch bedingte Unternehmenskrise erleiden als nicht pfadabhängige.

Aussagen über den kausalen Zusammenhang der beiden Merkmale sind ausgehend von den Ergebnissen nicht möglich. Es kann jedoch festgestellt werden, dass länderübergreifend eine signifikant positive Beziehung zwischen der Organisationalen Pfadabhängigkeit und der Anfälligkeit von Unternehmen für ethisch-moralisch bedingte Unternehmenskrisen besteht.

Folgende Überlegungen sollen bei der Interpretation dieser Ergebnisse beachtet werden:

- Pfadabhängigkeit ist ein langfristiges, evolutionäres Konstrukt, das durch eine wie auch immer geartete Fehlentscheidung initiiert wird. Je länger der Pfad sich dabei ausbildet, desto wahrscheinlicher entsteht eine (zunehmende) Diskrepanz zwischen interner und externer Entwicklung. In der heutigen globalisierten Welt entwickeln sich Gesellschaften extrem schnell, lernen und wachsen mit den Erfindungen und Ansprüchen der Zeit. Diese Entwicklung in Kombination mit persistenten, personell weitreichenden und thematisch konzentrierten Verhaltensstrukturen erhöht gemäß der Ergebnisse die Wahrscheinlichkeit des Aufeinanderprallens der internen und externen Perspektive in Form einer ethisch-moralisch bedingten Unternehmenskrise. Je länger die ethisch-moralisch unsauberen Strukturen bzw. Verhaltensmuster andauern und je mehr Personen diese bei gleichbleibenden Themen wie z.B. Korruption adaptieren, desto krisenanfälliger ist ein Unternehmen.

- Es zeigt sich zudem, dass sowohl die Werte der χ^2-Prüfstatistik als auch die der weiteren Korrelationsmaße (Cramer's V, Kontingenz-Koeffizient C, Phi) für die USA konsequent über den deutschen Werten liegen. Für Dow Jones 30-Unternehmen konnten für die drei Kriterien der Pfadabhängigkeit stark positiv signifikante Korrelationen mit der Anfälligkeit für ethisch-moralisch bedingte Unternehmenskrisen festgestellt werden. Die höheren amerikanischen Werte könnten dadurch begründet werden, dass in den USA entsprechende Gesetze und Verbote bereits früher umgesetzt wurden als in Deutschland. Als Beispiel kann der in den USA bereits seit 1989 existierende Whistleblower Protection Act angeführt werden. Hier wird ein Anreiz zum Anzeigen von Betrug oder sonstigen ethisch-moralisch bedenklichen Handlungen von Auftragnehmern der Bundesregierung gesetzt, indem Whistleblowern ca. 15 bis 25 Prozent der Schadenersatzsumme erhalten können („Bounty-Program"). Solche Prämien sollen dazu beitragen, eine Änderung der Unternehmenskultur herbeizuführen. John Kopchinski und fünf weitere Mitarbeiter der weltweit größten Pharmafirma Pfizer erhielten dank dieses Programms im Jahre 2009 mehr als 102 Mio. US Dollar dafür, dass sie illegale Praktiken ihres Unternehmens aufzeigten (vgl. Gehr, 2014). In Deutschland hatte die Bundesregierung 2008 eine Änderung des Bürgerlichen Gesetzbuches (BGB) mit dem Ziel eines verstärkten Whistleblower Schutzes angestrebt. Der neue Artikel ist in seinem vorgeschlagenen Wortlaut im Online-Appendix zu finden. Gerade in Deutschland wird jedoch die Grenze zum Denunziantentum besonders kritisch gesehen. Die Praxis zeigt jedoch, dass der Missbrauch anonymer Hinweisgebersysteme kaum relevant ist, vor allem wenn geregelt ist, dass Hinweisgeber in Missbrauchsfällen keinen Schutz genießen, wie es mittlerweile viele Codes of Conduct bereits vorsehen (vgl. Preedy, 2011: S. 63). Vor allem aufgrund der Widerstände innerhalb der CDU/ CSU wurde das Gesetzesvorhaben verworfen (vgl. Gehr, 2014). Für (DAX 30-) Unternehmen mit US-Börsennotierung schreibt jedoch der US Sarbanes-Oxley Act bereits seit 2002 für SOX-unterworfene internationale Unternehmen vor, dass eine vertrauliche und anonyme Meldemöglichkeit für Hinweise zu fragwürdigen Bilanzierungsmethoden einzurichten ist. Zusätzlich wurde der Schutz anonymer Hinweisge-

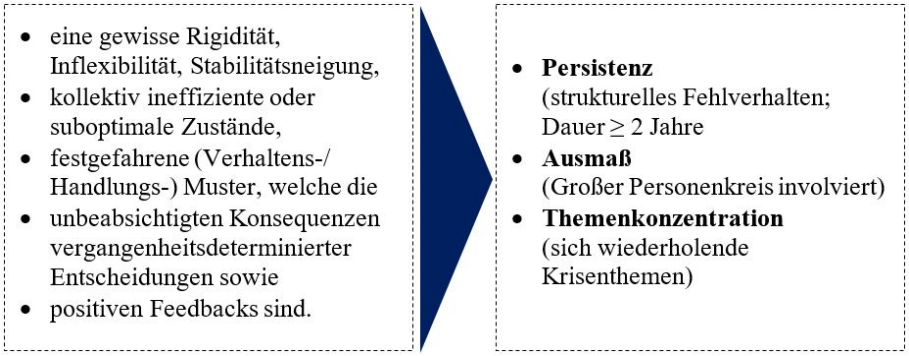

Abbildung 28: (Quelle: eigene Darstellung)

Tabelle 39: Vergleich der Pfadabhängigkeitsergebnisse von Deutschland und den USA

* DEU = Deutschland

	DEU*	USA	DEU*	USA	DEU*	USA	DEU*	USA
	χ^2-Wert		Cramer's V		Kontingenz Koeffizient C		Phi-Koeffizient	
Persistenz	16,676	19,33	0,541	0,572	0,476	0,497	0,541	0,572
Ausmaß	17,928	20,747	0,561	0,593	0,489	0,51	0,561	0,593
Themenkonzentration	9,529	16,978	0,409	0,536	0,378	0,473	0,409	0,536

ber vorgeschrieben. Infolgedessen mussten DAX 30-Konzerne mit damaliger US-Börsennotierung wie Siemens oder Infineon Whistleblower Hotlines einrichten (vgl. Mayer, 2013). Dies ist eines unter vielen Beispielen. Teils frühere und damit schnellere gesellschaftliche Entwicklungen oder Ansprüche an Unternehmen in den USA könnten als Begründung für eine relativ höhere Krisenanfälligkeit bei ähnlich persistenten Strukturen herangezogen werden. Besonders auffällig ist der χ^2-Werteunterschied beim Kriterium der Themenkonzentration. Hier liegt der amerikanische Wert extrem über dem deutschen Wert.

• Manchen Unternehmen gelingt es besser als anderen, eine solche ethisch-moralisch bedingte Unternehmenskrise zu nutzen, um mit persistenten und eventuell ineffizienten Verhaltensweisen zu brechen und so den Grundstein für ein künftiges krisenfreies Dasein zu legen. Diese Feststellung deckt sich auch mit den Erkenntnissen der in Abschnitt 2.3 vorgestellten Fallstudien von Tobias (2008) und Lüttel (2009). Basierend auf den hier gewonnenen Ergebnissen ist es jedoch schwierig darüber zu urteilen, ob es sich in Einzelfällen um ein Pfadbrechen oder lediglich um einen Pfadwechsel handelt. Bei der hier vorliegenden Arbeit zählt z.B. die MAN aufgrund der Schmiergeldaffäre 2009 zur Kategorie krisenanfälliger Unternehmen (vgl. sehr hoher zum Dekaden-Median relativer Google-Hits Wert). Dem Unternehmen ist es jedoch als Reaktion auf die ethisch-moralisch bedingte Krise gelungen, den Pfad bis dato zu verlassen. Die Krise kann hier als reinigende Bewegung, die dem Unternehmen eine Chance auf einen Neuanfang gegeben hat, gesehen werden. Für eine endgültige Unterscheidung und Beurteilung pfadbrechender und pfadwechselnder Unternehmen wird eine ergänzende Langzeitbetrachtung empfohlen.

3.4.8. Effektive Krisenpräventionsmaßnahmen

Grundsätzlich ist anzumerken, dass in den USA seit geraumer Zeit die haftungsrechtlichen Anforderungen nicht nur für börsennotierte Unternehmen durch Vorschriften wie die US Federal Sentencing Guidelines oder den Sarbanes-Oxley Act verschärft werden. Diese und weitere Regelungen fordern von Unternehmen zunehmend die Einführung von Verhaltenskodizes, Compliance-Programmen und Reviews, Anti-Betrugstrainings oder anonymen Hinweisgebersystemen. Solche Entwicklungen gehen in den letzten Jahren eher von den USA aus. Bei deutschen, bzw. allgemeiner gesprochen westeuropäischen, Unternehmen besteht im Vergleich zu nordamerikanischen Unternehmen bzgl. der Krisenprävention oft Nachholbedarf (vgl. Bussmann et al., 2010: S. 3). Die folgende Abbildung 29 stellt Maßnahmen dar, welche zur Prävention von Wirtschaftskriminalität eingesetzt werden, unterschieden nach der Art der Börsennotierung.

Es wird offensichtlich, dass vor allem US-börsennotierte

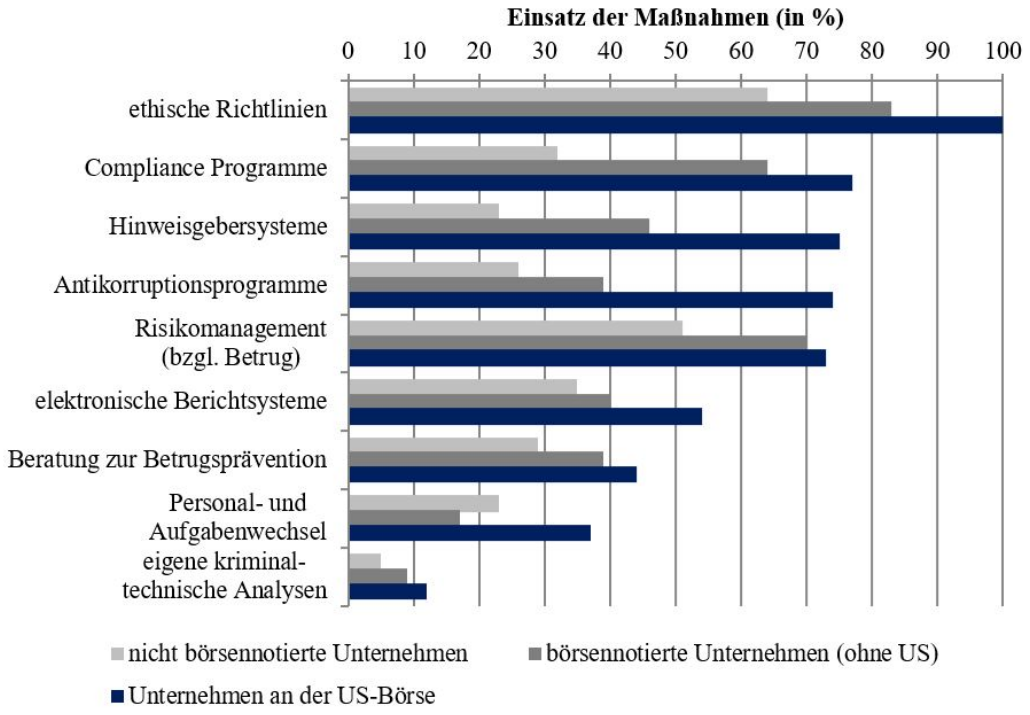

Abbildung 29: Maßnahmen zur Prävention von Wirtschaftskriminalität (Quelle: Bussmann et al., 2010, S. 13 f.)

Unternehmen sehr aktiv in der Umsetzung dieser Maßnahmen sind, was nicht zuletzt den zuvor vorgestellten rechtlichen Rahmenbedingungen geschuldet ist. Demzufolge verfügen 100 Prozent der Unternehmen an der US-Börse über Codes of Conduct/ Ethics und 77 Prozent sogar über Compliance-Programme. Dieser Kollektion von Maßnahmen bildet recht gut die hier getroffene Auswahl potenzieller Präventionsmaßnahmen ab. Die hier durchgeführten χ^2-Unabhängigkeitstests in der gemeinsamen Analyse der Länderdaten verdeutlichen, dass in Deutschland und den USA durchaus unterschiedliche Maßnahmen signifikant bzw. bei gemeinsamer Berechnung andere Maßnahmen signifikant sind (siehe Tabelle 40). Von Maßnahmen, die krisenvermeidend eingesetzt werden, wird eine signifikante negative Korrelation mit der Krisenanfälligkeit von Unternehmen erwartet werden.

Festgestellt werden konnte jedoch sowohl in der Einzelanalyse der Länder sowie bei der länderübergreifenden Betrachtung eine überwiegend (signifikante) positive Korrelation des Merkmals „Krisenanfälligkeit von Unternehmen“ und den untersuchten potenziellen Präventionsmaßnahmen. Bezüglich des Rotationsprinzips konnte für Deutschland, die USA und länderübergreifend eine negative, jedoch nicht signifikante, Korrelation beobachtet werden. Im Fall von Issues Management, Reputationsmanagement, Corporate Governance, CSR, Whistleblower Hotlines und Compliance-Systemen ist länderübergreifend eine signifikante positive Korrelation festzustellen. Ausgehend von den Ergebnissen des χ^2-Unabhängigkeitstests können keine Aussagen über eine mögliche Kausalität des Zusammenhangs gemacht werden. Maßnahmen wie ein Red-Flags System, regelmäßige Rotation, Integrity Management, unabhängige Krisenberater oder Risiko-/ Chancenmanagement sind länderübergreifend nicht signifikant. Es offenbaren sich hiermit einige Unterschiede in den Ergebnissen der einzelnen Länder im Vergleich zur gemeinsamen Betrachtung. Bei den Maßnahmen Corporate Governance, Whistleblower Hotlines und Compliance-Systemen liegt der Schluss nahe, dass es sich um Stichprobengrößeneffekte handelt. Diese Maßnahmen waren in der getrennten Betrachtung weder innerhalb der deutschen noch der amerikanischen Daten signifikant, sind es jedoch bei der länderübergreifenden Analyse. Weitere Forschung mit größeren Stichproben könnte hier Abhilfe schaffen.

Red-Flags Systeme und Integrity Management sind zwar in Deutschland signifikant, jedoch weder in den USA noch länderübergreifend. Es wird vermutet, dass zum einen in den USA und in Deutschland unterschiedliche Maßnahmen von Unternehmen favorisiert werden, dass gewissermaßen „Trends“ bestimmter Maßnahmen bestehen, welche nicht länderübergreifend, sondern oft nur landesweit gelten. Des Weiteren kann dies Folgen deutscher Krisenthemenschwerpunkte (Korruption, Spionage, Bestechung) widerspiegeln. Wie im Rahmen dieser Arbeit herausgestellt, steht Korruption bei deutschen Krisenfällen in der Mehrzahl der Fälle im Mittelpunkt. Die Vermutung des reaktiven Einsatzes von Maßnahmen hinnehmend erscheint es logisch, dass Unternehmen in Deutschland vorrangig korruptionspräventive Maßnahmen, z.B. ein Red-Flags System, einsetzen. In den USA beherrschen andere Krisenthemen (Finanzmarktmanipulation etc.) die Krisenlandschaft, was andere Maßnahmen

Tabelle 40: Vergleich der Ergebnisse bzgl. Präventionsmaßnahmen von Deutschland und den USA

D / USA Diese Maßnahmen waren in Deutschland (D) oder den USA (USA) signifikant
* signifikant zu einem 0,5-Prozent Signifikanzniveau; kritischer Wert: 7,88
** signifikant zu einem 2,5-Prozent Signifikanzniveau; kritischer Wert: 5,02
*** signifikant zu einem 5-Prozent Signifikanzniveau; kritischer Wert: 3,84

	χ^2-Wert	Cramer's V	Kontingenzkoeffizient C	Phi-Koeffizient
Issues Management^{D+USA}	10.762*	0.305	0.291	0.305
Reputationsmanagement^{D+USA}	8.579*	0.272	0.262	0.272
Corporate Governance	5.028**	0.208	0.204	0.208
CSR^{USA}	5.028**	0.208	0.204	0.208
Whistleblower Hotline	4.169***	0.19	0.186	0.19
Compliance-System	4.147***	0.189	0.186	0.189
Red-Flags System^{D}	2.519	0.147	0.146	0.147
Integrity Management^{D}	2.519	0.147	0.146	0.147
Unabhängiger Krisenberater	1.901	0.136	0.135	0.136
Rotationsprinzip	0.99	0.092	0.092	0.092(-)
Risiko-/ Chancenmanagement	0.838	0.085	0.085	0.085

in den Fokus rücken. Zudem legen die US-Richtlinien den amerikanischen Unternehmen andere Maßnahmen nahe. Darüber hinaus kann der bei Dow Jones 30-Unternehmen weniger stark ausgeprägten Einsatz z.B. von Red-Flags Systemen vermutlich auch auf die abweichende Bezeichnung eines solchen Systems zurückgeführt werden. Das wohl berühmteste Red-Flags System wurde von Siemens mitunter als Reaktion auf die Schmiergeldaffäre 2006 eingeführt bzw. ausgebaut. Im angelsächsischen Raum wird ein derartiges System möglicherweise anders bezeichnet.

Bezogen auf potenzielle Präventionsmaßnahmen für ethisch-moralisch bedingte Krisen konnte festgestellt werden, dass diese keineswegs dem Motto „mehr ist besser" folgen. Unternehmen, die viele bzw. alle untersuchten Maßnahmen durchführen, sind keineswegs garantierter Weise krisenfrei. Mehr Maßnahmen schützen nicht immer besser, sondern deuten oft auch auf eine krisenbehaftete Vergangenheit hin. Die in der länderübergreifenden Betrachtung beobachtete positive Korrelation von sechs Maßnahmen mit dem Kriterium der Krisenanfälligkeit legt den Schluss nahe, dass Unternehmen keine strategische Krisenprävention betreiben, sondern die Maßnahmen eher reaktiv einsetzen. Wer mehr Krisen zu verzeichnen hat, also krisenanfälliger ist, rüstet sich demzufolge mit eben diesen Maßnahmen und Mechanismen. Ist ein Unternehmen gerade durch eine Krise gegangen, stellt es Recherchen an, welche Maßnahmen zur Bekämpfung genau dieser Krisenursache herangezogen werden können und führt daraufhin z.B. eine die erlebte Krisenursache vermeidende Maßnahme ein. Dem vielfach in der Unternehmenskrisenliteratur unterstellten krisenpräventiven Nutzen der hier untersuchten Maßnahmen betonen die Ergebnisse hier den krisenverarbeitenden bzw. -vorbereitenden Charakter. Diesen reaktiven Charakter verdeutlicht auch ein von der BITKOM veröffentlichter Bericht im Bereich der Compliance-Programme. Die Compliance-Thematik erhielt dem Bericht zu Folge in Deutschland erst durch Aufsehen erregende Rechtsfälle wie dem Siemens-Fall hohe Priorität. Als Reaktion darauf sahen sich auch viele andere (in diesem Fall deutsche) Unternehmen im Zugzwang. So konnte der vermehrte Einsatz von Maßnahmen zur Wiederherstellung ihres Ansehens, um die Rechtsdurchsetzung aktiv zu fördern und ähnliche Rechtsverstöße zukünftig zu vermeiden, beobachtet werden (vgl. Bayrak et al., 2012: S. 6). Bezugnehmend auf den in Abschnitt zwei vorgestellten Crisis Management Life Cycle nach Coombs wäre zu erwarten gewesen, dass Unternehmen im Rahmen der Krisenprävention aktiv Maßnahmen ergreifen. Die Ergebnisse deuten jedoch eher darauf hin, dass sie auf der Stufe der Krisenvorbereitung agieren. Hier geht es vor allem um das Bewusstsein der Organisation, dass nicht alle Krisen vermieden werden können und es daher umso bedeutender ist, stets auf die nächste Unternehmenskrise vorbereitet zu sein (vgl. Gammelmark, 2011: S. 9).

Die Untersuchung zeigt des Weiteren, dass Unternehmen, egal ob krisenanfällig oder nicht, größtenteils Compliance-Systeme, Corporate Governance Strukturen, Risikomanagement sowie CSR implementieren, dies also kein relevantes Differenzierungsmerkmal oder gar effektiver Präventionsmechanismus zu sein scheint. Im Zuge weiterer Forschungsaktivitäten sollten daher weitere, hier nicht untersuchte Maßnahmen auf ihre Eignung zur Prävention hin beleuchtet werden.

Besonders die sehr großen im jeweiligen Index gelisteten Unternehmen betreiben viele Maßnahmen, sind also sehr

aktiv in Bezug auf die untersuchten Maßnahmen. Dies kann vor allem damit zusammenhängen, dass diese höhere Summen an Geld in vermeintliche Präventionsmaßnahmen stecken. Eine Vermutung ist des Weiteren, dass gerade die sehr großen Unternehmen eher einen „Legitimationsdruck“ am Markt und vor allem der Öffentlichkeit ggü. verspüren als die kleineren im Index gelisteten Unternehmen.

Es befinden sich vor allem Konsumenten nahe Unternehmen unter den aktiven Maßnahmen implementierenden Unternehmen. Diese scheinen sich der Öffentlichkeitsaufmerksamkeit bewusst zu sein und zeigen sich in ihrem Maßnahmen daher umso aktiver.

Es zeigt sich zudem, dass nahezu alle Unternehmen (DAX 30 und Dow Jones 30) Compliance-Programme, CSR und Corporate Governance Aktivitäten sowie Chancen-/ Risikomanagement betreiben, was diese Maßnahmen zu Hygienefaktoren qualifiziert. In den USA konnte des Weiteren das Vorhandensein einer Whistleblower Hotline als solche identifiziert werden. Diese Maßnahmen gleichen einer „organisationalen Modeerscheinung“, der nahezu alle Unternehmen Brachen übergreifend Folge leisten. Den Ergebnissen zur Folge weisen diese zwar keinen krisenpräventiven Charakter auf, jedoch könnte deren Abwesenheit durchaus Einfluss auf die Krisenanfälligkeit haben. Eine tiefgehende qualitative Analyse dieser Maßnahmen könnte daher wertvolle zusätzliche Erkenntnisse liefern. Vor allem am Beispiel von Reputationsmanagement wird deutlich, dass es nicht darauf ankommt, dass ein Unternehmen hierbei aktiv ist. Vielmehr ist die Wahrnehmung der Reputation am Markt bzw. der wesentlichen Zielgruppen ausschlaggebend. So betreiben Goldman Sachs zwar Reputationsmanagement, gehören aber neben JPMorgan Chase & Co. sowie der Bank of America zu den Unternehmen mit der schlechtesten Reputation in den USA (vgl. Merritt, 2014).

Auffällig ist darüber hinaus, dass der Einsatz unabhängiger Krisenberater in den USA populärer zu sein scheint als bei deutschen Unternehmen. In den USA konnte deren Einsatz in 14, in Deutschland nur in fünf Fällen beobachtet werden. Deutsche Unternehmen wie bspw. die Siemens AG, welche über vier Jahre hinweg durch Theo Waigel als obersten Anti-Korruptions-Aufseher im Auftrag der US-Behörden überwacht wurde (vgl. FOCUS, 2012), sind hierzulande die Ausnahme. US Aufsichtsbehörden wie die Securities and Exchange Commission (SEC) verpflichten Unternehmen des Öfteren des Hinzuziehens unabhängige Berater und Aufseher, welche die Verhaltensänderungen bzw. die Bemühungen darum überwachen sollen.

4. Diskussion

Der Diskussionsabschnitt gibt zu Beginn die mittels der fallbasierten Untersuchung gewonnenen eigenen Erkenntnisse wider und geht anschließend auf Implikationen für die Unternehmenspraxis ein. Abschließend wird auf bestehende Forschungslücken und Limitationen der vorliegenden Arbeit eingegangen.

4.1. Eigene Findings

Die vorliegende explorative Arbeit kann einige wertvolle Einsichten bezüglich ethisch-moralisch bedingter Unternehmenskrisen und deren potenzieller Prävention liefern.

Es hat sich gezeigt, dass ethisch-moralisch bedingte Krisen immer und überall, d.h. in jedem Unternehmen und zu jedem Zeitpunkt passieren können. Die untersuchten Krisen haben keine absolut gültigen und typischen Krisenstrukturen erkennen lassen. Zuverlässige Prognosen über das Auftreten oder den Verlauf dieser Krisen sind daher nicht pauschal möglich. Betrachtung von Einzelfällen, also Krisenmuster einzelner Unternehmen, können als Grundlage für Einschätzungen herangezogen werden. Ethisch-moralisch bedingte Krisen passieren dennoch nicht zufällig – vor dem Hintergrund der untersuchten Pfadabhängigkeitstheorie können sie zumeist als Ergebnis einer länger andauernden Auseinanderentwickelung von Innen (Unternehmen) und Außenperspektive (Gesellschaft) auftreten. Die Ergebnisse zeigen, dass Krisenanfälligkeit von Unternehmen mit den hier gewählten Kriterien der Pfadabhängigkeit korreliert. Maßnahmen zum Brechen oder Lösen pfadabhängiger Strukturen könnten demzufolge ein Ansatzpunkt der Prävention ethisch-moralisch bedingter Unternehmenskrisen bieten.

In Bezug auf potenzielle Präventionsmaßnahmen konnte zum einen festgestellt werden, dass in Deutschland und in den USA teils unterschiedliche Maßnahmen signifikant mit der Krisenanfälligkeit von Unternehmen korrelierten. Länderübergreifend korrelierten hierbei positiv die Maßnahmen Issues Management, Reputationsmanagement, Corporate Governance Strukturen, CSR, Whistleblower Hotlines sowie Compliance-Systeme. Zum anderen wurde durch die Ergebnisse gezeigt, dass eben diese Maßnahmen einen reaktiven Charakter aufweisen, d.h., dass Unternehmen diese eher als Reaktion auf eine ethisch-moralisch bedingte Unternehmenskrise implementieren als diese als strategische Präventionsmaßnahmen einzusetzen. Einige Maßnahmen konnten dabei als „Hygienefaktoren“ eingestuft werden, da nahezu alle Unternehmen, losgelöst vom Merkmal der Krisenanfälligkeit, diese durchführen. So zeigten sich Unternehmen wie Volkswagen und Daimler etc. in Deutschland bzw. Chevron und Coca-Cola usw. in den USA besonders aktiv durch den Einsatz fast aller Maßnahmen. Es hat sich jedoch herausgestellt, dass mehr Maßnahmen nicht automatisch einen besseren Krisenschutz bedeuten.

Eine hier behandelte Unternehmenskrise wurde in Abschnitt 2.1.1 als „Preventable Crises“ eingeordnet. Es liegt also das Verständnis zugrunde, dass eine Vermeidung der Krisen grundsätzlich möglich ist. Wie soeben beschreiben, lassen die hier untersuchten elf Krisenpräventionsmaßnahmen einen eher reaktiven Einsatz vermuten. Jedoch weisen die Ergebnisse darauf hin, dass die Erkenntnis des positiv signifikanten Zusammenhangs zwischen Krisenanfälligkeit von Unternehmen und Pfadabhängigkeit auch zur Prävention ethisch-moralisch bedingter Krisen genutzt werden kann. Maßnahmen, die dabei helfen, dass selbstverstärkende Rückkopplungseffekte nicht automatisch in einem negativen Lock-In münden und das Unternehmen nicht den

Bezug zur Außenperspektive verlieren, könnten durchaus zur Krisenprävention herangezogen werden. Mögliche Maßnahmen bauen auf die Fähigkeit des Unternehmens sich selbst zu reflektieren auf. Auf das Gedankengut der Theorie Organisationaler Pfadabhängigkeit aufbauend, sollten Unternehmen eine ethisch-moralische Unternehmenskrise trotz aller Wirkungen und Nachwirkungen als Chance sehen und wahrnehmen, um sich selbst und die vermutlich festgefahrenen Strukturen und Verhaltensweisen anzuerkennen und zu ändern, d.h. den Pfad zu brechen. Ohne die Fähigkeit der Selbstreflexion kann eine Unternehmenskrise zwar den einen Pfad brechen, jedoch ist die Wahrscheinlichkeit, dass ein neuer Pfad begangen wird, wahrscheinlich.

4.2. Managerial Implications

Die untersuchten Krisenfälle erlauben einerseits ein tieferes Verständnis ethisch-moralisch bedingter Unternehmenskrisen. Andererseits verbleiben bzgl. der Prognose und damit der Prävention solcher Krisenfälle einige offene Fragen. Festgestellt werden konnte, dass pfadabhängige Entwicklungen eine signifikante Rolle bei der Krisenanfälligkeit von Unternehmen spielen. Im Umkehrschluss könnten Mechanismen zur Eindämmung von Pfadabhängigkeit bzw. der Entwicklung in diese einen Ansatzpunkt für strategische Krisenprävention bieten.

Gesellschaften entwickeln sich weiter, oft sogar in einem Tempo, mit welchem die interne Entwicklung eines Unternehmens nicht Schritt halten kann. Unter anderem aufgrund der zunehmenden Diskrepanz der Entwicklungsstände folgt der möglicherweise unumgängliche Zusammenprall – die Unternehmenskrise. Den Ergebnissen zufolge besteht eine signifikante positive Beziehung zwischen der Krisenanfälligkeit von Unternehmen und den drei hier untersuchten Kriterien der Pfadabhängigkeit. Davon ausgehend können Maßnahmen zur Vermeidung pfadabhängiger Prozesse, d.h. im Kern selbstverstärkender Rückkopplungseffekte, auch ein krisenpräventiver Charakter unterstellt werden. Eine bewusste Konzentration auf die Reflexion und Anerkennung der Implikationen der Organisationalen Pfadabhängigkeitstheorie eröffnet Unternehmen eine neue, sich nicht ausschließlich auf die Unternehmenskrise per se konzentrierende, sondern tiefer ansetzende Art der Krisenprävention. Unternehmen, denen es gelingt, pfadabhängige Strukturen bzw. deren Kernmechanismen (selbstverstärkende Rückkopplungseffekte) zu vermeiden, sind gemäß den Ergebnissen dieser Arbeit erfolgreicher in der Krisenprävention als Unternehmen, welche reaktiv eine der elf hier untersuchten Maßnahmen durchführen.

Grundsätzlich ist darauf hinzuweisen, dass Maßnahmen, welche die Außenperspektive mit der Innenperspektive des Unternehmens abgleichen und ggf. das Vornehmen entsprechender Anpassungen veranlassen, vielversprechend im Einsatz der Vermeidung pfadabhängiger Muster sind. Ein Unternehmen sollte sich zu jedem Zeitpunkt selbst reflektieren und, wenn vorhanden, seine „Scheuklappen“ zugunsten einer generellen Aufgeschlossenheit ggü. neuester gesellschaftlicher Entwicklungen abwerfen.

Im Folgenden sollen daher qualitative Ansatzpunkte für ein Vermeiden bzw. Entgegenwirken pfadabhängiger Entwicklungen und damit Hinweise für eine effektive Prävention ethisch-moralisch bedingter Unternehmenskrisen gegeben werden. Unternehmen können in Anlehnung an Sydow et al. (2009), wie in Abschnitt 2.2.2 dargestellt, in der Prävention bei der Vermeidung der Etablierung selbstverstärkender Rückkopplungseffekte ansetzen. Aus dieser Annahme ergeben sich vier Ansatzpunkte möglicher Präventionsmaßnahmen:

1. Lerneffekte
2. Komplementäre Effekte
3. Koordinationseffekte
4. Erwartungseffekte

Ziel aller möglichen pfadabhängigkeitsvermeidenden Maßnahmen ist eine Erweiterung des potenziell verengten Entscheidungs- und Handlungsspielraums. Es wird im Folgenden eine generelle qualitative Intuition möglicher Maßnahmen, die an eben diesen vier Ansatzpunkten ansetzen, gegeben. Zudem werden die hier untersuchten elf potenziellen Präventionsmaßnahmen auf ihre Eignung zum Einsatz als Instrument der Pfadabhängigkeitsvermeidung überprüft. Hinter dieser argumentativen Zuordnung der elf Maßnahmen verbirgt sich das Verständnis, dass, selbst wenn Unternehmen die Maßnahmen eher reaktiv einsetzen, diese zur Vermeidung künftiger ethisch-moralisch bedingter Unternehmenskrisen beitragen können, indem sie potenzielle pfadabhängige Strukturen unterbinden.

4.2.1. Lerneffekte

Unternehmen sollten Strukturen schaffen, die neue Ideen zulassen und sogar fördern. Der Mechanismus, dass einmal angewöhnte, ethisch-moralisch bedenkliche Strukturen und Handlungsweisen sich selbst in die Organisation einbrennen, muss aufgebrochen werden. Eine wichtige Rolle könnte hier auch neuen Organisationsmitgliedern zukommen, welche frische Impulse in das vermeidlich festgefahrene, pfadabhängige System einbringen. Dass gewisse Ausführungsweisen von Aufgaben etabliert sind, sollte nicht als Faktum hingenommen werden, sondern immer wieder neu hinterfragt werden. Diese Impulse müssen vom System bzw. der Organisation gezielt gefördert werden. Einer aufgeschlossenen Unternehmenskultur kommt hier sicherlich eine wichtige Rolle zu.

Das *Rotationsprinzip* kann hier einen sinnvollen Einsatz finden. Indem Aufgaben und Positionen in regelmäßigen zeitlichen Abständen gewechselt werden, gewöhnen sich Organisationsmitglieder an das Erlernen und Einbringen neuer Aufgaben und Ideen. Die Gefahr des Etablierens bestimmter Strukturen und Verhaltensweisen wird somit reduziert. Die Ergebnisse der Untersuchung haben angedeutet, dass auch das Prinzip der Rotation eine zwar nicht signifikante, aber negative Korrelation mit der Krisenanfälligkeit aufweist. Die Idee der auch in der Literatur zu Unternehmenskrisen häufig diskutierten Rotation ist, dass die Opportunitätskosten

ethisch-moralisch unsauberer Handlungen durch die regelmäßige Versetzung von Mitarbeitern in korruptionsgefährdeten Positionen in andere Positionen oder Abteilungen gesteigert werden können (vgl. Dietz, 1998: S. 87). In einer solchen rotierenden Arbeitssituation müssen sich immer neue Partner miteinander arrangieren, wodurch dem Aufbau langfristiger ethisch-moralisch bedenklicher Beziehungen vorgebeugt werden kann. Ein erhöhtes Entdeckungsrisiko ist ein Nebenprodukt dieser Maßnahme (vgl. Schmidt, 2004: S. 23). Der Einsatz eines *unabhängigen Krisenberaters* kann eine ähnliche Wirkung aufweisen. Ein externer Krisenberater bringt Benchmark-Ansätze ins Unternehmen, hinterfragt dort vorherrschende Strukturen und Verhaltensweisen kritisch und weist auf eventuelle Gefahren durch ein potenzielles Festfahren der Unternehmung hin. Beide Maßnahmen können zu einer Erweiterung des Handlungsspielraums führen. Abgesehen von den hier untersuchten elf Maßnahmen könnten Anreizsystem, interne Ideenmärkte etc. implementiert werden, die neue Ideen fördern.

4.2.2. Komplementäre Effekte

Hier gilt es, die Vorteile der Synergien, welche durch die Interaktion zweier oder mehrerer separater aber interkorrelierter Ressourcen, Regeln oder Verhaltensweisen erreicht werden, zu begrenzen.

Corporate Governance sowie *Compliance-Programme, Risiko-/ Chancenmanagement* und *Integrity Management* könnten hier potenziell als Maßnahmen der Prävention pfadabhängiger Strukturen verstanden werden. Im Grundsatz dienen sie dem Ziel der Eindämmung von Economies of Scope. *Risiko-/ Chancenmanagement* könnte die Rolle der Evaluation von Risiken und Präzisierung gefährdeter interkorrelierter Ressourcen, Organisationsmitglieder, Regeln oder Verhaltensweisen zukommen. Die Ergebnisse und Empfehlungen dieser ständigen Analyse könnten sich in Aktionen im Rahmen von *Corporate Governance* sowie *Compliance-Programmen* konkretisieren. Der Kommunikation klarer Unternehmensleitlinien und die Bestrafung bei Nichteinhaltung kommt dabei eine wichtige Rolle zu. *Integrity Management* könnte all diese Maßnahmen als ganzheitliches Konzept aufeinander abstimmen und koordinieren.

4.2.3. Koordinationseffekte

Durch die Vorteilhaftigkeit konformen Verhaltens resultiert dieser Prozess in einem gewissen festgefahrenen Verhalten der Organisationsmitglieder und damit der Organisation als Ganze. An vielen Stellen wirken sich Economies of Scale positiv aus, da sie u.a. die Koordinationskosten senken. In Bezug auf pfadabhängige Prozesse zeigen sich diese Effekte jedoch von ihrer „negativen Seite“. Je mehr Organisationsmitglieder ein bestimmtes Verhaltensmuster ausführen, desto attraktiver scheint dieses Verhaltensmuster. In Folge dessen führen es auch weitere Organisationsmitglieder durch. Der Druck zur Konformität steigt (vgl. Sydow et al., 2009: S. 699). Um dieser Entwicklung vorzubeugen, muss das Unternehmen zuerst erkennen, bei welchen (wichtigen) Verhaltensweisen oder Strukturen es sich um ungünstiger Weise Festgefahrene handelt. Dieser Erkenntnis muss das aktive Aufbrechen des Pfads folgen – die Vorteilhaftigkeit dieses Verhaltens bzw. dieser Struktur darf keine Ausschließlichkeit bleiben.

Das Einrichten eines *Red-Flags Systems* dient vor allem der unternehmensweiten Analyse für Betrug, Korruption etc. gefährdeter Bereiche (vgl. Landeskriminalamt NRW, 2007: S. 7) . Die Ergebnisse dieser Analyse könnten eine Indikation des Verbreitungsgrades bzw. Ausmaßes bestimmter (festgefahrener) Strukturen und Handlungsweisen geben. Zudem sollten bzgl. der Analyseergebnisse kritische Werte und Alarmniveaus sowie entsprechende Aktionsroutinen definiert werden. Eine Whistleblower Hotline könnte im Wirkungsfeld der Koordinationseffekte eingesetzt werden. Hier gilt es den Druck zur Konformität aufzulösen. Auch wenn die Masse der Organisationsmitglieder eine bestimmte Struktur bzw. eine Verhaltensweise lebt, so muss dies nicht „richtig“ sein. Um Hinterfragungen Raum zu geben und kritisch Denkenden ein Sprachrohr zu verleihen, kann eine Whistleblower Hotline helfen.

4.2.4. Adaptive Erwartungseffekte

Organisationsmitglieder adaptieren bestimmte Routinen und Verhaltensweisen und führen diese kontinuierlich in der Erwartung aus, dass „die anderen“ dasselbe tun. Die Angst des Outsider-Werdens treibt viele Organisationsmitglieder hierzu. Die Meinung „der anderen“ darf an dieser Stelle keine Black Box bleiben, es darf kein blindes Ausführen getrieben durch eine ggf. falsche Erwartungshaltung über die Meinung anderer Organisationsmitglieder stattfinden. Offene Gesprächsrunden, gesteigerte Transparenz etc. könnten hier Ansatzpunkte für die Unternehmenspraxis sein. Auch hier könnten entsprechend *Whistleblower Hotlines* zum Einsatz kommen, um eine anonyme Meinungsäußerung zu gewährleisten. *Reputationsmanagement*, im Sinne einer Kommunikation ins Innere des Unternehmens, könnte dabei die Funktion zukommen, die Erwartungen „der anderen“ greifbarer zu machen. Es sollte versucht werden, Erwartungen und Präferenzen zu kommunizieren, um eine blinde Anpassung an die nicht sichtbaren Präferenzen „der anderen“ zu vermeiden.

Issues Management und CSR-Aktivitäten konnten keinem dieser vier Ansatzpunkte nach Sydow et al. (2009) eindeutig zugeordnet werden. Andere Interpretationen bzw. Zuordnungen sind nicht ausgeschlossen. Über die in Anlehnung an die Autoren vorgestellten Ansatzpunkte selbstverstärkender Rückkopplungseffekte hinaus können folgende qualitativ beschriebene Maßnahmen Anregungen für Unternehmen zur Vermeidung pfadabhängiger Entwicklungen sein:

4.2.5. Integration negativen Feedbacks

Positives Feedback zählt zu den Grundmechanismen der Pfadabhängigkeit. Wird Fehlverhalten nicht bestraft, sondern im Gegenteil, gereicht sogar noch zu positivem Feedback, wird ein Festfahren eben dieses Verhaltens sehr wahrscheinlich. Bringt es mithilfe von Schmiergeldern erkauftes Geschäft mehr Umsatz als ein sauber zustande gekommenes

Geschäft, so wird in Zukunft mit hoher Wahrscheinlichkeit wieder Schmiergeld eingesetzt, dass sich dieses Verhalten offensichtlich lohnt. An dieser Stelle sollten Unternehmen zur Vermeidung einer Pfadabhängigkeit und damit zur Prävention ethisch-moralisch bedingter Krisen verstärkt negatives Feedback in allen Wertschöpfungsstufen integrieren, sei es in Form eines Kunden orientierten Beschwerdemanagements oder interner Möglichkeiten der Äußerung negativen Feedbacks. Die Erkenntnisse aus dieser Feedbackschleife sollten unbedingt von der Organisation reflektiert und als Chance frühzeitiger Krisenprävention gesehen werden. Denn eines haben alle Krisenfälle gezeigt: Ethisch-moralisch unsauberes Verhalten wird früher oder später immer bestraft und ist für Unternehmen häufig sehr teuer.

4.2.6. Transparenz

Hiermit ist sowohl eine Transparenz nach innen als auch nach außen gemeint. Eine Organisation sollte Transparenz nach außen fördern, denn nur dies ermöglicht, dass auch hier wertvolles (negatives sowie positives) Feedback ans Unternehmen zurückgetragen wird und kann auf längere Sicht das Auseinanderentwickeln der Innen- und Außenperspektive ermöglichen. Transparenz nach innen kann sich bspw. durch klare Kommunikation der Leitlinien und der Konsequenzen bei Nichteinhaltung oder auch durch einen offensiven Umgang bei vermuteten ethisch-moralisch unsauberer Strukturen oder Handlungen äußern. Die Gesellschaft verlangt zunehmend nach Transparenz, diese Entwicklung gilt es für Unternehmen ernst zu nehmen.

4.2.7. Prinzip der Lernenden Organisation

Eine „Lernende Organisation“ findet sich in der Literatur sehr unterschiedlich definiert. Gemein ist aber den meisten Definitionen, dass es sich um eine Organisation handelt, welche „nicht so sehr auf Vorschriften [...], sondern auf Ziele, Zwecke oder Leitgedanken ausgerichtet [ist]“ (Brodtrick, 1997: S. 204), und vor allem über die Gabe der Selbstreflexion verfügt. Auf ethisch-moralisch bedingte Krisen übertragen bedeutet dies, dass ein Unternehmen jeden Tag die Wahl hat, etwas besser zu machen, kritisch mit den eigenen Strukturen oder Verhaltensmustern zu sein und diese zu ändern. Aus geschehenen Krisen kann und sollte gelernt werden, sie sollten als Chance für einen Neuanfang jenseits der alten Strukturen gesehen und genutzt werden.

4.2.8. Funktionstrennung

Das Risiko ethisch-moralischer Fehltritte könnte des Weiteren vermindert werden, indem die Kombination bestimmter Funktionen und Entscheidungskompetenzen bei einer Person oder Stelle vermieden werden. Die Bereiche Entscheidung, Ausführung, Überprüfung, Berichterstattung und Abrechnung sollten getrennt werden (vgl. Schmidt, 2004: S. 23).

4.2.9. Vier-Augen-Prinzip

Diesem Prinzip zur Folge sollten gewisse organisationsrelevante Tätigkeiten nicht von einer Person alleine ausgeführt, sondern von mindestens einer weiteren Person mit „beaufsichtigt“ werden. Dieses Prinzip könnte z.B. in Korruptionsfällen implementiert werden. Ab einem bestimmten Auftragswert könnten Aufträge nicht von einem Organisationsmitglied alleine vergeben werden (vgl. Schmidt, 2004: S. 23).

Neben den hier aufgezeigten Möglichkeiten zur Prävention ethisch-moralisch bedingter Krisen bestehen viele weitere Ansätze. Wie sich Krisen in ihrer Art und Ausprägung verändern, so evolutionär bzw. adaptiv sollten auch deren Präventionsmaßnahmen sein. Jedoch sollte sowohl dem aktuellen als auch dem zukünftigen Krisenpräventionsverständnis die Einsicht zugrunde liegen, dass neben dem reinen Umsetzen einer Maßnahme auch weiche Faktoren wie die allgemeine Unternehmenskultur einen wesentlichen Einfluss auf die Effektivität haben (vgl. Schmidt, 2004: S. 23).

4.3. Forschungslücken und weitere Forschungsanregungen

Wie in Abschnitt 2.3 aufgezeigt, besteht zwar in den einzelnen Forschungsbereichen, dem der Unternehmenskrisen und dem Organisationaler Pfadabhängigkeit, eine große Masse verfügbarer und hochqualitativer Literatur. Jedoch ist die integrierte Forschung, welche versucht, einen Zusammenhang zwischen beiden Forschungsthemen zu finden und wertvolle Ableitungen für die Unternehmenspraxis zu gewinnen, noch nicht etabliert. Diese Arbeit ist ein Beitrag zur gemeinsamen Betrachtung beider Forschungsgebiete. Es verbleiben weiterhin offene Fragen, welche zum Inhalt zukünftiger Forschungsaktivitäten gemacht werden sollten. Hier einige spannende Ansatzpunkte für weitere Forschung, welche sich ausgehend von dieser Arbeit ergeben haben:

- Sowohl innerhalb des DAX 30 als auch des Dow Jones 30 konnten krisenfreie Unternehmen identifiziert werden. An dieser Stelle wäre es interessant genauer zu beleuchten, wie sich diese Unternehmen von stark krisengeprägten Unternehmen unterscheiden. Woran liegt es, dass Unternehmen wie z.B. Altana keine (bzw. weniger intensive) ethisch-moralisch bedingte Unternehmenskrisen erleiden? Das ist das Bild des Datensatzes. Vermutet wird, dass manche Unternehmen wie Siemens oder Daimler öffentlichkeitsnäher wahrgenommen werden, da sie für Endkunden produzieren und sie Öffentlichkeitswahrnehmung bzw. Aufmerksamkeit hier höher ist als bei Unternehmen, die hauptsächlich im B2B-Bereich agieren. Weniger öffentlichkeitsnahe Unternehmen haben demnach eine Art Wettbewerbsvorteil, da sie offensichtlich nicht weniger oft an strukturellen Problemen im Sinne einer Pfadabhängigkeit leiden, aber dies aufgrund der geringeren öffentlichen Aufmerksamkeit nicht zwangsläufig in eine ethisch-moralisch bedingte Unternehmenskrise führen muss.
- Die vorliegende Arbeit lebt von ihrem stark explorativen Charakter. Ergänzend wären weitere

auch quantitative Untersuchungen zum Thema Pfadabhängigkeit in Verbindung mit Unternehmenskrisen denkbar. Zukünftige Untersuchungen sollten zudem auf einer größeren Stichprobe basieren, um hier mitunter vorliegenden Stichprobengrößeneffekten entgegenzuwirken.

- Die Arbeit bietet bereits einen Vergleich zweier Länder bzw. deren Unternehmen. Allerdings sind sich Deutschland und die USA wirtschaftlich und gesellschaftlich betrachtet sehr nahe. Hier könnte ein Vergleich mit anderen Kultur-/ oder Wirtschaftskreisen wie bspw. dem asiatischen Raum spannende Erkenntnisse liefern.
- Bezüglich der untersuchten Maßnahmen wird eine weitere qualitative Untersuchung vorgeschlagen. Diese Arbeit hat sich vor allem mit einer „Bestandsaufnahme" der jeweiligen Aktivitäten bzw. Präventionsmaßnahmen beschäftigt. Eine ergänzende qualitative Betrachtung, welche sich mit der Güte der jeweiligen durchgeführten Aktivitäten beschäftigt, könnte weitere wertvolle Erkenntnisse liefern. So legt ein von Pricewaterhouse Coopers veröffentlichter Bericht nahe, dass die (Qualität der) Unternehmenskultur einen erheblichen Einfluss auf die Akzeptanz und in Konsequenz auf die Effektivität bestimmter Krisenpräventionsmaßnahmen hat. Die Autoren der Studien kommen zu dem Schluss, dass „der gute Wille allein [nicht genügt]" – Präventionsmaßnahmen sind keine reinen Organisationsaufgaben. Alle Maßnahmen bedürfen einer hohen Akzeptanz und einer positiven Unternehmenskultur, um ihre präventive Wirkung entfalten zu können (vgl. Bussmann et al., 2010: S. 4). Eine Untersuchung vor diesem Hintergrund würde eine sehr sinnvolle Erweiterung dieser Arbeit darstellen.

4.4. Limitationen der vorliegenden Arbeit

Limitationen der vorliegenden Arbeit beziehen sich vor allem auf zwei Hauptbereiche:

(1) Identifikation relevanter ethisch-moralisch bedingter Krisenfälle

(2) Kodierung und Auswertung der Krisenfälle im zeitlichen Verlauf Bei der Identifikation relevanter ethisch-moralisch bedingter Krisenfälle soll vor allem auf die folgenden möglichen Einschränkungen hingewiesen werden:

- Die Suchbegriffe sind sehr subjektiv und in Folge dessen reagiert das Ergebnis der Google Hits sehr sensibel auf unterschiedliche Suchbegriffe.
- Tendenziell sind „alte Fälle" (Krisenfälle vor dem Jahr 2000) aufgrund ihrer geringen online-Verfügbarkeit bzw. Digitalisierung unterrepräsentiert. Dieser Einschränkung wurde jedoch versucht mit der Verhältnisbildung einzelner Krisen-Hits mit dem Dekaden-Median entgegenzuwirken.
- Es konnte im Rahmen dieser Arbeit nur ein bestimmter Zeitraum betrachtet werden. Pfadabhängigkeit ist ein sehr langfristiges Konstrukt. Entwicklungen dahin sind so gut wie nie endgültig abgeschlossen und damit auch nur schwer beobachtbar.
- Unternehmensprozesse sind teils sehr undurchsichtig. In der Deutung der Fälle ist es oft schwierig, das auslösende Ereignis, also das small event, zu identifizieren. Unternehmen werden nie alle Informationen, selbst bei hohen Transparenzansprüchen nach einer Krise, offenlegen, vor allem nicht nachträglich. So wird das Bild der Pfadabhängigkeit im Einzelfall nie lückenlos aufgeklärt werden können. Hierzu fehlt es an sensiblen Unternehmensinformationen.
- Verflechtungen mit anderen Unternehmen sind oft undurchsichtig – vor allem bei Übernahmen von Unternehmen kommt dies zum Tragen. Hier stellt sich früher oder später die Frage, wer (welches Unternehmen), welche Aktionen oder Entscheidungen zu verantworten hat(te). Auch in solchen Sonderfällen wird dies vermutlich nie nachvollzogen werden können.

Bezüglich der Kodierung bzw. Auswertung der Krisenfälle im zeitlichen Verlauf könnten folgende Aspekte eine Einschränkung bedeuten:

- Die Liste möglicher Präventionsmaßnahmen (Whistleblower-Hotline etc.) erhebt keinen Anspruch auf Vollständigkeit. Es wurde versucht, die zumeist praktiziertesten Maßnahmen aufzugreifen, ebenso wie die primären Empfehlungen aus der Forschungsliteratur.
- Die Kausalität mancher Maßnahmen ist nicht immer klar: Wurde bspw. eine Compliance-Abteilung als Reaktion auf eine Krise eingerichtet oder handelte es sich um eine Organisations-Entscheidung? Kam eine weitere Krise nicht, eben weil diese Compliance-Abteilung nun agiert und präventiv gewirkt hat. Um dieser Limitation entgegenzuwirken, wurde sowohl bei der Analyse der Pfadabhängigkeit als auch bei der Bewertung effektiver Präventionsmaßnahmen von einer Betrachtung je Krise auf eine Betrachtung je Unternehmen übergegangen. Die gewonnenen Beobachtungen und Implikationen dienen daher vielmehr der Erweiterung des Verständnisses rund um ethisch-moralisch bedingte Unternehmenskrisen als der exakten Dokumentation der (Krisenpräventions-) Maßnahmen von Unternehmen.

• Der χ^2-Unabhängigkeitstest ist ein relativ einfaches, aber dennoch weit verbreitetes Testverfahren. Problematisch für die Prüfstatistik sind kleine Werte pro Zelle (Faustregel < 3). Bei großen Stichproben der χ^2-Unabhängigkeitstest sehr zuverlässig. Im vorliegenden Fall kamen mitunter auch Zellwerte von Null vor. Dies ist bei der Interpretation der Ergebnisse zu beachten.

5. Konklusion

Die vorliegende explorative Arbeit hat wertvolle Erkenntnisse zum Verständnis ethisch-moralisch bedingter Unternehmenskrisen vermittelt. Die gemeinsame Betrachtung ethisch-moralisch bedingter Unternehmenskrisen und der Theorie Organisationaler Pfadabhängigkeit liefert Ansatzpunkte für eine effektive Prävention dieser Krisen. Pfadabhängige Strukturen vermeidende Maßnahmen wie z.B. ein hohes Maß an Transparenz oder das Prinzip der Lernenden Organisation können nicht nur die Dominanz bestimmter Verhaltensweisen oder Strukturen über andere verhindern. Ausgehend von dem festgestellten signifikanten Zusammenhang zwischen einer Pfadabhängigkeit und der Krisenanfälligkeit von Unternehmen weisen Pfadabhängigkeit vermeidende Maßnahmen ebenso einen krisenpräventiven Charakter auf.

Die empirische Untersuchung sowie ihre Ergebnisse und Implikationen sollen darüber hinaus als Anreiz für zukünftige Forschungsanstrengungen gesehen werden. Hier untersuchte Unternehmenskrisen werden auch in Zukunft immer wieder zu öffentlichen Diskussionen führen und die Zeitungen Tag um Tag füllen. Bemühungen um das Verständnis solcher ethisch-moralisch bedingten Krisen sind daher sowohl für die Forschung als auch für die Unternehmenspraxis von essentieller Bedeutung.

Literatur

Altmeier, L. Kleine Schmierereien unter Freunden. Hamburg:ZEIT ONLINE GmbH, 2012. URL http://www.zeit.de/politik/2012-11/korruption-bestechung-politik-alltag. Abgerufen: 21.05.2014.

Andrews, C. und LeBlanc, B. Fraud Hotlines: Don't Miss That Call. *Journal of Accountancy*, 216(2):32–35, 2013.

Ansari, S. und Garud, R. Inter-generational transitions in socio-technical systems: The case of mobile communications. *Research Policy*, 38(2):382–392, 2009.

Antonelli, C. The economics of path-dependence in industrial organization. *International Journal of Industrial Organization*, 15(6):643–675, 1997.

Argenti, P. *Corporate Communication*. Mcgraw Hill Book Co, 2007.

Arnold, H.-P. *Die Krise früh genug erkennen*. Deutscher Drucker, 14 edition, 2010.

Arthur, W. Competing Technologies, Increasing Returns, and Lock-In by Historical Events. *The Economic Journal*, 99:116–131, 1989.

Arthur, W. Positive Feedbacks in the Economy. *Scientific American*, pages 80–85, 1990.

Arthur, W. *Increasing returns and path dependence in the economy*. University of Michigan Press, 1994.

Asprion, P. Design der empirischen Analyse. In *Funktionstrennung in ERP-Systemen*, pages 183–213. Springer, 2013.

Bayrak, M.-A., Harder, B., Herzog, C., Horn, M., Kemperman, P., Lechleitner, J., Mayer, C., Schramm, A., Streit, T., Vondenhoff-Mertens, R., und Witteler, S. Leitfaden Compliance - Rechtliche Anforderungen an ITK-Unternehmen. BITKOM Bundesverband Informationswirtschaft Telekommunikation und neue Medien e. V., 2012.

Beyer, J. Pfadabhängigkeit ist nicht gleich Pfadabhängigkeit! Wider den impliziten Konservatismus eines gängigen Konzepts. *Zeitschrift für Soziologie*, 2005.

Borchardt, A. und Göthlich, S. Erkenntnisgewinnung durch Fallstudien. In Albers, S., Klapper, D., Konradt, U., Walter, A., und Wolf, J., editors, *Methodik der empirischen Forschung*, number 33-48. Gabler, 2007.

Brodtrick, V. Die Lernende Organisation. In *Qualitäts-und erfolgsorientiertes Verwaltungsmanagement: aktuelle Tendenzen und Entwürfe*. Duncker & Humbolt, Berlin, 1997.

Bromley, D. *Reputation, image and impression management*. John Wiley & Sons, 1993.

Bussmann, K., Krieg, O., Nestler, C., Salvenmoser, S., Schroth, A., Theile, A., und Trunk, D. Compliance und Unternehmenskultur–Zur aktuellen Situation in deutschen Großunternehmen. Pricewaterhouse Coopers, 2010.

Carter, S. The Interaction of Top Management Group, Stakeholder, and Situational Factors on Certain Corporate Reputation Management Activities. *Journal of Management Studies*, 43(5):1145–1176, 2006.

Coombs, W. *Ongoing Crisis Communication: Planning, Managing, and Responding*. SAGE, 3 edition, 2007a.

Coombs, W. Protecting Organization Reputations During a Crisis: The Development and Application of Situational Crisis Communication Theory. *Corporate Reputation Review*, 10(1057):163–176, 2007b.

Coombs, W. und Holladay, S. Communication and Attributions in a Crisis: An Experiment Study in Crisis Communication. *Journal of Public Relations Research*, 8(4):279–295, 1996.

Coombs, W. und Holladay, S. Helping Crisis Managers Protect Reputational Assets: Initial Tests of the Situational Crisis Communication Theory. *Management Communication Quarterly*, 16(2):165–186, 2002.

David, P. Clio and the Economics of QWERTY. *American Economic Review*, 75:332–337, 1985.

Dietz, M. *Korruption: eine institutionenökonomische Analyse*. Spitz, Berlin, 1998.

Dobusch, L. und Kapeller, J. Breaking New Paths: Theory and Method in Path Dependence Research. *Schmalenbach Business Review*, 65(3):288–311, 2013.

Dobusch, L. und Schüßler, E. Theorizing path dependence: a review of positive feedback mechanisms in technology markets, regional clusters, and organizations. *Industrial & Corporate Change*, 22(3):617–647, 2013.

Dubrovski, D. Management mistakes as causes of corporate crises: Managerial implications for countries in transition. *Total Quality Management & Business Excellence*, 20(1):39–59, 2009.

Economist. Mortgage-related bank fines - Payback time for subprime. The Economist Newspaper Limited, 2013. URL http://www.economist.com/news/finance-and-economics/21588386-payback-time-subprime. Abgerufen: 03.06.2014.

Eisenhardt, K. Building theories from case study research. *Academy of management review*, 14(4):532–550, 1989.

Fleck, T. Issue Management - Krisenprävention aber auch Chancenkommunikation. ILTIS GmbH, 2014. URL http://4managers.de/management/themen/issue-management/. Abgerufen: 07.06.2014.

FOCUS. Ex-Finanzminister lobt Kampf gegen Korruption - Theo Waigel beendet Kontrollfunktion bei Siemens. FOCUS Online, 2012. URL http://www.focus.de/finanzen/news/unternehmen/ex-finanzminister-lobt-kampf-gegen-korruption-theo-waigel-beendet-kontrollfunktion-bei-siemens_aid_850130.html. Abgerufen: 06.07.2014.

Fombrun, C. und Shanley, M. What's in a name? Reputation building and corporate strategy. *Academy of Management Journal*, 33(2):233–258, 1990.

Freitag, M. und Katzensteiner, T. Compliance - Terror der Tugend. manager magazin new media GmbH, 2012. URL http://www.manager-magazin.de/magazin/artikel/a-842925.html. Abgerufen: 21.04.2014.

Gammelmark, G. Crisis Communication: Planning or No Planning? An assessment of the importance of crisis planning. Master's thesis, Aarhus School of Business, Aarhus University, 2011.

Gehr, W. Whistleblowing International. Whistleblowing Austria, 2014. URL http://www.whistleblowing.at/Whistleblowing_Austria/International.html. Abgerufen: 18.06.2014.

Hartmann, J. und Tauber, A. Wie Dax-Konzerne den Schmiergeldsumpf überwinden. Axel Springer SE, 2012. URL http://www.welt.de/wirtschaft/article13827833/Wie-Dax-Konzerne-den-Schmiergeldsumpf-ueberwinden.html. Abgerufen: 07.04.2014.

Heath, R. und Palenchar, M. *Strategic issues management: Organizations and public policy challenges*. SAGE, 2008.

Henning, M., Stam, E., und Wenting, R. Path Dependence Research in Regional Economic Development: Cacophony or Knowledge Accumulation? *Regional Studies*, 47(8):1348–1362, 2013.

Heybrock, H. *Praxisratgeber Compliance - Compliance Managementsysteme in Funktionsbereichen des Unternehmens*. Rainer Hampp, München, Mering, 2012.

Institut für Krisenforschung. Krisennavigator Kiel, 2014. URL https://www.krisennavigator.de/Krisenforschungsinstitut.krisenforschungsinstitut.0.html. Abgerufen: 21.05.2014.

Jossé, G. und Ziegelmeyer, A.-C. Ein Instrument zur Krisenerkennung für den mittelständischen Einzelhandel. *CONTROLLER Magazin*, 5:80–87, 2013.

Kahn, W., Barton, M., und Fellows, S. Organizational Crises And The Disturbance Of Relational Systems. *Academy of Management Review*, 38(3):377–396, 2013.

Kay, N. Rerun the tape of history and QWERTY always wins. *Research Policy*, 42(6-7):1175–1185, 2013.

Kennedy, J. F. History Quotes - John F. Kennedy Quotes, 2014. URL http://historyquotes.org/john-f-kennedy-quotes/. Abgerufen: 06.04.2014.

Khalil, E. Lock-in institutions and efficiency. *Journal of Economic Behavior & Organization*, 88:27–36, 2013.

Knauß, F. Der Skandal ist überall. Düsseldorf: Handelsblatt GmbH, 2012. URL http://www.wiwo.de/erfolg/management/krisenkommunikation-allzeit-droht-der-skandal/6927292-2.html. Abgerufen: 18.04.2014.

Koch, J. und Rothmann, W. Zur Überwindung organisationaler Pfadabhängigkeit durch Coaching: Ansatzpunkte zur Diagnose und Intervention. In *Die Organisation in Supervision und Coaching*, pages 81–94. VS Verlag für Sozialwissenschaften, 2010.

Korschun, D., Bhattacharya, C., und Swain, S. Corporate Social Responsibility, Customer Orientation, and the Job Performance of Frontline Employees. *Journal of Marketing*, 78(3):20–37, 2014.

Kotler, P. und Lee, N. *Corporate Social Responsibility: Doing the Most Good for Your Company and Your Cause*. John Wiley & Sons, 2005.

Küttner, C. Simmons &Simmons betreibt die ThyssenKrupp Whistleblower Hotline. RWS Verlag Kommunikationsforum GmbH, 2007. URL http://www.rws-verlag.de/hauptnavigation/aktuell/news-detail/period/1191189600/2681999/archived/browse/4/select/newsticker_kanzleien/article/340/Simmons-Simmons-b

etreibt-die-ThyssenKrupp-Whistleblower-Hotline-1.html. Abgerufen: 05.07.2014.

Landeskriminalamt NRW. Korruptionsprävention für Banken und Wirtschaftsunternehmen, 2007.

Leyendecker, H. Siemens: Korruptionsaffäre. Süddeutsche.de, 2011. URL http://www.sueddeutsche.de/wirtschaft/siemens-korruptionsaffaere-das-ist-wie-bei-der-mafia-1.1046507. Abgerufen: 08.04.2014.

LÜBECK. Typen von Unternehmenskrisen. LÜBECK Steuerberater Rechtsanwälte GbR, 2014. URL http://www.luebeckonline.com/mustervertraege/insolvenz/typen-von-unternehmenskrisen.html. Abgerufen: 12.05.2014.

Lüttel, J. *Pfadabhängigkeit in strategischen Entscheidungsprozessen - Die Entwicklung der Bankgesellschaft Berlin in den Jahren 1994 bis 2001*. Peter Lang, Frankfurt am Main, 2009.

Mahoney, J. Path dependence in historical sociology. *Theory and society*, 29 (4):507–548, 2000.

Mayer, E. Whistleblowing – Fluch und Segen von Hinweisgeber-Systemen. Berlin: Berufsverband der Compliance Manager (BCM) e.V., 2013. URL http://www.bvdcm.de/news/whistleblowing-fluch-und-segen-von-hinweisgeber-systemen. Abgerufen: 21.06.2014.

McGreal, P. Corporate Compliance Survey. *Business Lawyer*, 69(1):107–132, 2013.

Merritt, J. Reputation Rankings: Facts and Lessons from the Harris Poll Reputation Quotient. ReputationManagement.com, 2014. URL http://www.reputationmanagement.com/reputation-rankings-facts-lessons-harris-poll-reputation-quotient/. Abgerufen: 26.06.2014.

Mitroff, I., Shriwastava, P., und Udwadia, F. Effective Crisis Management. *Academy of Management Executive*, 1(4):283–292, 1987.

Mohr, H. Evolutionäre Ethik. *Evolutionary Ethics*, 21(2):231–242, 2010.

Odrakiewicz, P. und Odrakiewicz, D. Integrity Management and Anti-Corruption Actions in an Organizational Context. *Global Management Journal*, 6(1):65–73, 2014.

Ortmann, G. *Formen der Produktion: Organisation und Rekursivität*. VS Verlag für Sozialwissenschaften, Wiesbaden, 1995.

Pelofsky, J. Corporate monitor may be in Wal-Mart's checkout lane. Thomson Reuters, 2012. URL http://www.reuters.com/article/2012/05/11/us-walmart-mexico-monitors-idUSBRE84A13620120511. Abgerufen: 21.06.2014.

Pierson, P. Increasing returns, path dependence, and the study of politics. *American political science review*, 94(2):251–267, 2000.

Polley, N. und Kroner, S. Unternehmen in der Krise? - Handlungsempfehlungen zur Rolle des Aufsichtsrats. Der Aufsichtsrat (05), 2012.

Preedy, K. Nachhaltigkeit ist entscheidend. *Personalmagazin - Management, Recht und Organisation*, 8:63–64, 2011.

Prorokowski, L. und Prorokowski, H. Organisation of compliance across financial institutions. *Journal of Investment Compliance (Emerald Group)*, 15(1):65–76, 2014.

Ricken, F. *Allgemeine Ethik*. C.H. Beck, Stuttgart, 1983.

Rüsen, T. Verständnis und Modelle von Unternehmenskrisen. In *Krisen und Krisenmanagement in Familienunternehmen*, pages 43–77. Gabler, 2009.

Schmidt, S. *Korruption in Unternehmen - Typologie und Prävention*. PhD thesis, Fachhochschule Pforzheim, 2004.

Schreyögg, G., Sydow, J., und Koch, J. *Organisatorische Pfade - von der Pfadabhängigkeit zur Pfadkreation? Strategische Prozesse und Pfade*. Gabler, Wiesbaden, 2003.

Schreyögg, G., Sydow, J., und Holtmann, P. How history matters in organisations: The case of path dependence. *Management & Organizational History*, 6(1):81–100, 2011.

Schwaiger, M. Cross Tabulations. Management, 2013.

Shrivastava, P. und Mitroff, I. Strategic Management of Corporate Crises. *Columbia Journal of World Business*, 22(1):5–11, 1987.

Simmie, J. Path Dependence and New Path Creation in Renewable Energy Technologies. *European Planning Studies*, 20(5):729–731, 2012a.

Simmie, J. Path Dependence and New Technological Path Creation in the Danish Wind Power Industry. *European Planning Studies*, 20(5):753–772, 2012b.

Sohn, Y. und Lariscy, R. Understanding Reputational Crisis: Definition, Properties, and Consequences. *Journal of Public Relations Research*, 26(1): 23–43, 2014.

Steinbuch, A. Der ständige Kampf gegen Korruption - Mittelständler müssen besonders aufpassen. WirtschaftsWoche, 2013. URL http://www.wiwo.de/unternehmen/mittelstand/wirtschaftskriminalitaet-im-mittelstand-der-staendige-kampf-gegen-korruption/8716138.html. Abgerufen: 07.04.2014.

Sydow, J. Organisationale Pfade: Wie Geschichte zwischen Organisationen Bedeutung erlangt. In *Die Ökonomie der Organisation – die Organisation der Ökonomie*, pages 15–31. VS Verlag für Sozialwissenschaften, 2010.

Sydow, J., Schreyögg, G., und Koch, J. Organizational Path Dependence: Opening the Black Box. *Academy of Management Review*, 34(4):689–709, 2009.

Tellis, W. Results of a case study on information technology at a university. *The qualitative report*, 3(4):76, 1997.

Thelen, K. Historical Institutionalism in Comparative Politics. *Annual Review of Political Science*, 2(1):369–404, 1999.

Tobias, R. Krise durch Pfadabhängigkeit?–Sanierung durch Pfadbruch! Die Beherrschung pfadförmiger Krisendynamiken, dargestellt an einem Praxisfall. *Krisen-, Sanierungs- und Insolvenzberatung: Wirtschaft, Recht, Steuern 4*, 2008.

Topper, B. und Lagadec, P. Fractal Crises - A New Path for Crisis Theory and Management. *Journal of Contingencies & Crisis Management*, 21(1):4–16, 2013.

Toutenburg, H. und Heumann, C. *Induktive Statistik - Eine Einführung mit R und SPSS*. Springer, Berlin, 2000.

Transparency. Korruptionsindizes. Transparency International Deutschland e.V., 2014. URL http://www.transparency.de/Korruptionsindizes.1015.0.html. Abgerufen: 21.05.2014.

TV-Novosti. Banks paid $100 bn in US fines for 'pre-crisis mistakes'. Autonomous Nonprofit Organization "TV-Novosti", 2014. URL http://rt.com/business/banks-pay-us-penalties-313/. Abgerufen: 03.06.2014.

Vergne, J. QWERTY is dead - long live path dependence. *Research Policy*, 42 (6-7):1191–1194, 2013.

Wolf, K. *Risikomanagement im Kontext der wertorientierten Unternehmensführung*. Deutscher Universitäts-Verlag, 2003.

Wolke, T. *Risikomanagement*. Oldenbourg, 2008.

Wöltje, J. Frühindikatoren für Unternehmenskrisen. In *Betriebswirtschaftliche Formelsammlung*, pages 243–288. Rudolf Haufe, Freiburg, 2011.

Yin, R. *Case Study Research: Design and Methods*. SAGE, Beverly Hills, 1984.

Junior Management Science

journal homepage: www.jums.academy

Potenziale von Batteriespeichersystemen zur Regelenergiebereitstellung

Johannes Weindl
Technische Universität München

Abstract

Die Energiewirtschaft befindet sich zur Zeit in einem Umbruch. Der wachsende Anteil der fluktuierenden Stromproduktion von Erneuerbaren Energien und die zunehmende Verdrängung konventioneller Kraftwerke gefährdet die Stabilität des Stromnetzes. Deshalb müssen neue Lösungen zur Sicherstellung einer verlässlichen Energieversorgung gefunden werden. Batteriespeichersysteme können ein bedeutender Baustein für den Erfolg dieser Energiewende sein. Vom technischen Standpunkt sind sie besonders dazu geeignet Primärregelleistung bereitzustellen. Folglich wird das entsprechende Geschäftsmodell durch eine Kapitalwertberechnung für ein fiktives Lithium-Ionen und ein Bleisäure Batteriespeichersystem, basierend auf Marktdaten und neu eingeführten Vorschriften, untersucht. Leider verhindern die gegenwärtigen Marktbedingungen in Deutschland einen profitablen Betrieb beider Speichertechnologien, wobei das Bleisäure Batteriespeichersystem ein vielversprechenderes Ergebnis liefert.

Keywords: Batteriespeicher, Energiewende, Primärregelleistung, Regelenergie, Wirtschaftlichkeit

1. Die Energiewende und ihre Folgen

1.1. Umbruch in der Energiewirtschaft

In der Vergangenheit wurde die Energieversorgung von Wasserkraftwerken und Dampfkraftwerken, nuklearer und fossiler Art, zuverlässig bereitgestellt und war voll kontrollierbar. In jüngster Zeit befindet sich die Energieerzeugung aus Erneuerbaren Energien (EE) auf dem Vormarsch. So gab es bereits im Jahr 2012 Zeiten, in denen die Energie aus Windenergieanlagen (WEA) und Photovoltaik(PV)-Anlagen rund 50% des Strombedarfs abgedeckt haben.[1] Im Szenariorahmen zum Netzentwicklungsplan 2025 geht die Bundesnetzagentur von einem Ausbau der gemeinsamen installierten Leistung von WEA und PV-Anlagen auf 167 GW bis zum Jahr 2035 aus. Das entspricht mehr als einer Verdopplung des Referenzwertes aus dem Jahr 2013 von rund 71 GW. Im gleichen Zeitraum sinkt die Erzeugungsleistung aus konventionellen Anlagen von 101 GW auf 77,5 GW, während die Jahreshöchstlast nahezu stagniert und im Jahr 2035 bei 84 GW liegen soll.[2]

Im westlichen Stromnetz[3] Dänemarks wurde bereits 2007 zu Spitzenzeiten mehr Strom in WEA erzeugt als verbraucht werden konnte. Der Anteil der Windenergie an der gesamten Stromerzeugung betrug damals schon 24%. Dänemark kann daher als ein Schaubild für die Struktur und Eigenschaften für zukünftige Energienetze anderer Länder gesehen werden: Ein großer Teil der Energieerzeugung kann nicht mehr aktiv kontrolliert werden, da er aus dargebotsabhängigen Quellen, wie Wind und Sonne, stammt.[4] Die Netzbetreiber stellt das vor einige Herausforderungen hinsichtlich der Zuverlässigkeit der Energieversorgung, der steigenden Netzeinspeisung im Niedervoltbereich und der Kosten für die Netzinfrastruktur. [5] Während die Einspeisung aus PV-Anlagen noch einem Tageszyklus folgt, existiert bei WEA keine solche Regelmäßigkeit. Um Gefährdungen der Systemstabilität durch deren fluktuierende Stromeinspeisung zu verringern, gehört es daher zum Aufgabenbereich der deutschen Übertragungsnetzbetreiber (ÜNB), täglich eine Windprognose für den folgenden Tag zu erstellen.[6]

1.2. Speicher für die Energiewende

Während die konventionellen Kraftwerke durch den Ausbau der EE immer weiter zurückgedrängt werden, wächst

[1] Boje et al. (2014b, S. 7290-7291)
[2] Bundesnetzagentur (2014, S. II)
[3] Das dänische Stromnetz ist zweigeteilt. Der westliche Teil verhält sich synchron mit dem kontinentaleuropäischen Netz , der östliche Teil mit dem Nordischen Netz
[4] Divya und Østergaard (2009, S. 512, 517)
[5] Koller et al. (2015, S. 128)
[6] Verband der Netzbetreiber – VDN – e.V. beim VDEW (2007, S. 16–17)

DOI: http://dx.doi.org/10.5282/jums/v1i1pp275-300

gleichzeitig der Bedarf an der Regelleistung, die sie bisher zum größten Teil selbst bereitgestellt hatten.[7] So errechneten Hudson et al., dass eine installierte Windenergieleistung von 100 Megawatt (MW), einen Bedarf an Regelenergie von 6% der Nennleistung zur Folge hat. Dieser Prozentwert sinkt allerdings mit zunehmender installierter Windenergieleistung.[8] Insbesondere die negative Regelleistung, bei der Energie aus dem Netz entnommen wird, wird in Zukunft eine immer wichtigere Rolle spielen, da es durch den weiteren Ausbau der EE immer häufiger ein Energieüberangebot im Netz geben wird.[9]

Dabei können auch die EE selber einen Beitrag zur Bereitstellung von Regelleistung bringen: Durch Herunterregeln ihres Arbeitspunktes können WEA negative Regelleistung bereitstellen. Auch das Erbringen positiver Regelenergie ist möglich, sofern die Anlagen unterhalb ihrer maximalen Erzeugungsleistung betrieben werden. Das ist heute jedoch noch nicht Realität, da Anreizsysteme bisher nur eine maximale Erzeugung begünstigen und die Betriebsverfahren der Regelleistung erst entsprechend angepasst werden müssten.[10] Betreiber von Windturbinen in Dänemark sind hingegen schon heute zur Erbringung von Regelleistung verpflichtet: Der Arbeitspunkt der WEA muss beliebig zwischen 20% und 100% der Nennleistung einstellbar sein. Im Falle einer Windflaute muss natürlich keine Regelleistung erbracht werden können.[11] In Zukunft werden aus Sicht der Übertragungsnetze Windparks wie heutige konventionelle Kraftwerke funktionieren und auch in der Lage sein, Regelleistung zu erbringen. Möglich wird dies durch das Einbeziehen von Batteriespeichern, die zusammen mit den WEA ein sogenanntes Virtuelles Kraftwerk bilden.[12]

Bisher werden nur Pumpspeicher als großformatige Energiespeicher dazu eingesetzt, gespeicherte Energie im regulären Energiehandel zu vermarkten. Batteriespeicher wären zwar unabhängig von der Geographie und auch schnell realisierbar, sind aber aus Kostengründen noch nicht konkurrenzfähig im Vergleich mit Pumpspeichern.[13] So kostet das Einspeichern von 1 kWh Strom in einen Batteriespeicher weit mehr als 10 Cent,[14] während der Börsenstrompreis im Juni 2015 bei nur 3 ct/kWh lag.[15] Doch um die Energiewende in Deutschland verwirklichen zu können, werden neben den Pumpspeichern auch andere großformatige Speichertechnologien nötig sein. So sieht die Bundesregierung in ihrem Energiekonzept vom 28.09.2010 den Ausbau von Energiespeicherkapazitäten sowie die Entwicklung und Förderung von Speichertechnologien, wie Batteriespeicher, als eines der Handlungsfeldern auf dem Weg in das Zeitalter der EE.[16] Das Fraunhofer-Institut sieht vor allem die Flexibilität im Stromversorgungssystem, sowohl auf Erzeuger- als auch auf Verbraucherseite als einen zentralen Baustein für das Gelingen der Energiewende. Jedoch halten die Forscher bis zu einem EE-Anteil von 60% und bei ausreichender Flexibilisierung selbst bei einem EE-Anteil von 90% keinen zusätzlichen Ausbau von Stromspeichern für erforderlich. Allerdings sieht das Fraunhofer-Institut in der Bereitstellung von Systemdienstleistungen, zu denen unter anderem auch die Regelenergie gehört, eine Anwendung, in der Speichersysteme, auf Grund zukünftig steigender Anforderungen, einen nennenswerten Beitrag leisten können.[17] Auch Fuchs et al. sehen die Flexibilität des zukünftigen Stromnetzes als einen Faktor für den Erfolg der Energiewende. Dabei sind für sie auch Energiespeicher ein Baustein für die Flexibilität. Allerdings vertreten sie die Auffassung, dass sich ab 2030 nur noch Speichersysteme lohnen werden, die neben einer Systemdienstleistung mindestens eine weitere Aufgabe erfüllen. Für die Übergangszeit bis 2030 halten die Autoren den Einsatz von Batteriespeichern für sinnvoll.[18]

1.3. Aufbau und Zielsetzung der Bachelorarbeit

Kapitel 1.2 verdeutlicht, dass Energiespeicher eine gewichtige Rolle bei der Bewältigung der deutschen Energiewende spielen werden. In dieser Arbeit soll daher untersucht werden, inwiefern sich Batteriespeicher für die Bereitstellung von Regelenergie eignen und daran anknüpfend das entsprechende Geschäftsmodell auf seine Rentabilität überprüft werden. Dazu werden zunächst in Kapitel 2 das Konzept der Regelenergie erläutert, die drei Arten von Regelleistung vorgestellt und die Fokussierung der weiteren Arbeit auf die Primärregelleistung (PRL) begründet. Bevor in Kapitel 4 auf Batteriespeichersysteme, ihre Komponenten sowie ihre Eignung für die Erbringung von PRL eingegangen wird, erfolgt in Kapitel 3 eine Einführung in die dafür grundlegenden Batterietechnologien. In Kapitel 5 erfolgt dann eine Wirtschaftlichkeitsberechnung für einen fiktiven Batteriespeicher der PRL anbietet. Dazu werden reale Daten aus dem Regelenergiemarkt ausgewertet sowie Informationen von Betreibern und Herstellern von Batteriesystemen einbezogen. Im abschließenden 6. Kapitel wird ein Ausblick auf die zukünftigen Entwicklungen des Regelenergiemarktes und der Batteriespeichertechnologien gegeben.

2. Regelenergie

In einem Energieversorgungsnetz müssen sich die Stromerzeugung und der Stromverbrauch stets ausgleichen. Ein Ungleichgewicht zwischen Produktion und Verbrauch führt zu einer Abweichung der Netzfrequenz von ihrem Sollwert

[7] Stöckl (2014, S. 79)
[8] Hudson et al. (2001, S. 4)
[9] Fuchs et al. (2012, S. 52)
[10] Rivero et al. (2001, S. 337–338)
[11] Elkraft System and Eltra Regulation (2004, S. 4–6)
[12] Swierczynski et al. (2013, S. 487)
[13] Borsche et al. (2013, S. 1)
[14] Kempkens (2014b)
[15] Internationales Wirtschaftsforum Regenerative Energien (2015)
[16] Deutscher Bundestag (2010, S. 12–15)
[17] Fraunhofer-Institut für Windenergie und Energiesystemtechnik et al. (2014, S. 8)
[18] Fuchs et al. (2012, S. 1, 55–56)

und damit zum Absinken der elektrischen Versorgungsqualität.[19] Kann die Netzfrequenz nicht in einem bestimmten Bereich gehalten werden, führt dies zu Instabilitäten im Netzbetrieb und kann schlimmstenfalls einen Stromausfall verursachen.[20] Bei den Ursachen für die Frequenzabweichung unterscheidet man zum einen die stochastische Lastabweichung, die jederzeit, an beliebigen Orten mit verschiedenen Vorzeichen auftreten kann. Dabei kann es sich um den Ausfall eines Kraftwerks handeln oder das Zu- und Abschalten von Lasten auf der Verbraucherseite, aber auch Ereignisse, die sich über mehrere Stunden hinziehen können, wie die Veränderung der Außentemperatur. Auf der anderen Seite gibt es auch deterministische Lastabweichungen, ausgelöst durch Änderungen im Fahrplan der Stromerzeugungseinheiten.[21] Die Regelung der Netzfrequenz betrifft das gesamte zusammenhängende Übertragungsnetz, da Frequenzabweichungen nicht örtlich begrenzt, sondern global im ganzen Verbundnetz auftreten.[22] Zur Frequenzregelung wird ein bestimmter Teil der Erzeugungskapazität bzw. der Last genutzt um aktive Leistung bereitzustellen bzw. aufzunehmen. [23] Diese Regelleistung wird bisher hauptsächlich von thermischen Kraftwerken und Wasserkraftwerken erbracht, die ein Stück unterhalb ihrer Nennkapazität betrieben werden, damit sie kurzfristig zusätzliche Leistung zur Verfügung stellen können, falls es zu einem Ausfall einer anderen Erzeugungseinheit kommt.[24]

2.1. Regulatorischer Rahmen der Regelenergie in Deutschland

Während die Regeln des Netzwerks der europäischen Übertragungsnetzbetreiber (ENTSO-E) den europäischen Rahmen für die Durchführung der Frequenzregelung bilden, sind in Deutschland die gesetzlichen Vorgaben der Stromnetzzugangsverordnung (StromNVZ) und des Energiewirtschaftsgesetzes (EnWG) maßgeblich. Die dazugehörigen Marktregeln werden von der Bundesnetzagentur festgelegt und bei Bedarf angepasst. Technische Regularien, beispielsweise für die Präqualifikation als Anbieter von Regelleistung, sind dem Transmission Code der deutschen ÜNB zu entnehmen.[25]

Das "Operations Handbook" des ENTSO-E sieht für das kontinentaleuropäische Verbundnetz ein dreistufiges Regelleistungskonzept vor. Die einzelnen Stufen werden dabei nacheinander durchlaufen. Auf eine Abweichung der Netzfrequenz von ihrem Sollwert, 50 Hertz (Hz), reagiert zunächst die Primärregelleistung (PRL) innerhalb von Sekunden. Diese wird nach einigen Minuten von der Sekundärregelleistung (SRL) abgelöst. Anders als die PRL, die von allen Teilnetzen des Verbundes gemeinsam erbracht wird, erfolgt die Erbringung der SRL nur in der Regelzone, die für die Frequenzabweichung verantwortlich ist. Diese ist auch alleinig für die 3. Stufe des Regelleistungskonzepts, die sogenannte Minutenreserveleistung (MRL) zuständig. Sie ergänzt die SRL zunächst und ersetzt diese schließlich komplett.[26] Laut § 13 Abs. 1 EnWG sind die Übertragungsnetzbetreiber für die Zuverlässigkeit der Stromversorgung in ihrer Regelzone verantwortlich und sollen diese unter anderem durch den Einsatz von Regelenergie sicherstellen. Die Beschaffung der Regelenergie hat nach § 22 Abs. 2 EnWG über eine gemeinsam von den Übertragungsnetzbetreibern betriebene Internetplattform zu erfolgen. Das Ausschreibungsverfahren muss dabei transparent und diskriminierungsfrei sein. So sollen, soweit dies technisch möglich ist, für alle Anbieter die gleichen Anforderungen gelten. Insbesondere darf es laut § 22 Abs. 1 EnWG keine Bevorzugung von Unternehmen geben, die mit den ÜNBs verbunden oder assoziert sind.[27] Haben die Übertragungsnetzbetreiber zur Abrufung von Regelleistung früher noch direkten Einfluss auf den Betrieb von Kraftwerken genommen, erfolgt also heutzutage die Beschaffung auf freiwilliger Basis über einen Markt.[28] Der Handel von Regelleistung stellt für die ÜNBs und die Anbieter der Regelleistung eine Win-Win-Situation dar, führt er doch zu einem Interessensausgleich der beiden Parteien: Während die ÜNB die Zuverlässigkeit des Netzes bei möglichst geringen Kosten sicherstellen wollen, versuchen die Anbieter ihren Gewinn zu maximieren.[29]

2.2. Der deutsche Regelleistungsmarkt im europäischen Kontext

Am deutschen Markt sind, Stand Februar 2014, 14 Anbieter für PRL und 20 Anbieter für SRL präqualifiziert. Dabei handelt es sich überwiegend um Betreiber von Großkraftwerken. Hingegen sind unter den 36 präqualifizierten Anbietern von MRL auch Stromerzeuger mit kleineren Produktionsanlagen und Großabnehmer von Strom. Für eine Einordnung des deutschen Regelleistungsmarkts in den europäischen Kontext ist die Marktgröße der einzelnen Regelleistungsarten interessant: Die PRL ist im Netzverbund der ENTSO-E so ausgelegt, dass der Ausfall der größten Erzeugungseinheit im System ausgeglichen werden kann. Im kontinentaleuropäischen Netz sind zu diesen sogenannten Referenzereignis 3.000 Megawatt (MW) bereitzustellen. Da die PRL solidarisch geleistet wird, kann diese Beschaffungsmenge auf die einzelnen Regelzonen, je nach deren Anteil an der gesamten Stromerzeugung, aufgeteilt werden. Für Deutschland ergab sich so 2014 eine Ausschreibungsgröße von 568 MW. Die Verfahren zur Dimensionierung von SRL und MRL sind europaweit weniger einheitlich. In Deutschland erfolgt die Bemessung quartalsweise über ein wahrscheinlichkeitsbasiertes Verfahren, das eine Beschaffungsmenge toleriert,

[19] Oudalov et al. (2007, S. 1260)
[20] Boje et al. (2014b, S. 7290)
[21] Weissbach und Welfonder (2009, S. 27)
[22] Lachs und Sutanto (1995, S. 701)
[23] Rivero et al. (2001, S. 334)
[24] Oudalov et al. (2006, S. 2208)
[25] Consentec GmbH (2014, S. 1–2)
[26] ENTSO-E - European Network of Transmission System Operators for Electricity (2009, S. P1-2)
[27] Deutscher Bundestag (2005)
[28] Oudalov et al. (2007, S. 1260)
[29] Rebours et al. (2007b, S. 358)

die in einigen wenigen Stunden des Jahres nicht ausreichend ist. Im ersten Quartal 2014 ergaben sich so für positive und negative SRL 2.042 MW bzw. 1.969 MW und entsprechend bei MRL 2.472 MW bzw. 2.838 MW.[30]

Der sogenannte "reserve indicator" (RI) ist ein Maß für die Größe eines Regelleistungsmarktes. Er ist das Ergebnis der Division von durchschnittlich vorgehaltener Regelleisung im System und dem Durchschnitt der Energieerzeugung pro Stunde.[31] Ein kleiner RI bedeutet somit, dass eine Regelzone im Verhältnis zu ihrer Energieproduktion wenig Regelleistung vorhält. Man könnte dies somit als ein Zeichen für eine hohe intrinsische Systemstabilität und eine zuverlässige Stromversorgung sehen, da in jener Regelzone wohl wenig Regelleistung benötigt wird. Andererseits könnte ein kleines RI auch schlichtweg von weniger Sorgfalt bei der Sicherung der Netzstabilität herrühren. Eine weitere Möglichkeit Regelleistungsmärkte zu vergleichen ist der sogenannte "cost indicator" (CI), der aus dem Quotient der Jahresgesamtkosten für Regelleistung und der Jahresgesamtkosten der erzeugten Energie nach Großmarktpreisen ermittelt wird. So lag der CI für PRL in Deutschland in den Jahren 2004 - 2005 bei ca. 0,5% und damit auf einem mit anderen europäischen Staaten vergleichbaren Niveau. Der deutsche CI für die SRL betrug im gleichen Zeitraum ca. 1,5%[32]

2.3. Technische Grundlagen der Primärregelleistung

Die Primärregelleistung hat die Aufgabe die Netzfrequenz, nach einer Abweichung von ihrem Sollwert, zu stabilisieren, aber nicht auf ihren Nennwert zurückzuführen. Die abgegebene bzw. aufgenommene Leistung verhält sich dabei proportional zur Frequenzabweichung und wird von allen Regelzonen eines zusammenhängenden Netzes solidarisch erbracht.[33] Zur Aktivierung der PRL ist keine zentrale Regeleinrichtung nötig, da sie allein durch die innerhalb des Verbundnetzes identische Netzfrequenz gesteuert wird. Die Regelung erfolgt daher dezentral bei den einzelnen Erbringern der PRL, beispielsweise durch Turbinendrehzahlregler bei Kraftwerken.[34] Die sogenannte Statik ist dabei ein Maß für die Sensitivität einer Erzeugungseinheit, die PRL bereitstellt, gegenüber Abweichungen von der Sollfrequenz: Je kleiner die Statik ist, desto stärker reagiert die betreffende Erzeugungseinheit mit einer Änderung der abgegebenen Leistung auf eine bestimmte Frequenzabweichung. Sie wird durch den Quotient aus der relativen Abweichung von der Sollfrequenz und der relativen Leistungsänderung gebildet.[35]

In einem bestimmten Bereich um den Sollwert der Netzfrequenz herum erfolgt keine Reaktion der PRL. Diese Unempfindlichkeit gegenüber einer Frequenzabweichung ist nur zu einem Teil beabsichtigt. Sie entsteht auch durch Ungenauigkeiten bei der Messung der Netzfrequenz bei den dezentralen Reglern der PRL.[36] Im Verbundnetz der ENTSO-E muss bis zu einer gemessenen Abweichung der Netzfrequenz von +/- 10 Millihertz (mHz) vom Sollwert keine Aktivierung der PRL erfolgen. Dieser Frequenzbereich wird als Totband bezeichnet. Die Frequenzmessung selbst muss dabei mindestens auf 10 mHz genau sein. Zusammengerechnet kommt es so spätestens bei einer realen Frequenzabweichung von +/- 20 mHz zur Aktivierung der PRL. Bis zur vollen Abrufung der PRL, die bei einer Frequenzabweichung von +/- 200 mHz erfolgt, wächst die aktivierte Leistung linear mit der Frequenzabweichung. Die Aktivierung der PRL in Abhängigkeit von der Frequenzabweichung ist in Abbildung 1 dargestellt [37] Aus der Statik und der Frequenz, bei der die maximale PRL-Abrufung erfolgt, lässt sich somit die Leistung errechnen, die im normalen Betrieb vorgehalten werden muss, um die Anforderungen der PRL zu erfüllen.[38] Da es bei Batteriespeichern im Ruhezustand der PRL keine Leistungsabgabe gibt und somit die gesamte Nennleistung für die Erbringung von PRL zur Verfügung steht, kann die Statik hier folglich als ein Maß für die bei einer bestimmten Frequenz absolut zu erbringende Leistung gesehen werden.

Da PRL solidarisch geleistet wird, haben Länder bzw. Regelzonen, die Teil eines großen Verbundnetzes, wie dem kontinental-europäischen, sind, in der Regel einen niedrigeren RI-Wert für die PRL, als solche, die keine oder nur wenige Verbindungen zu anderen Regelzonen unterhalten. In ersteren verteilt sich nämlich die aus dem Referenzereignis abgeleitete Beschaffungsmenge der PRL auf mehrere Regelzonen.[39] Wie stark die einzelnen Regelzonen auf eine Frequenzabweichung reagieren wird durch deren Frequenzcharakteristik Lamda beschrieben. Sie errechnet sich aus der Abweichung des tatsächlichen vom fahrplanmäßigen Leistungsaustausch mit anderen Regelzonen.[40]

2.4. Sekundärregelleistung

Bei der SRL wird die Abrufung durch einen automatischen Leistungs-Frequenzregler beim ÜNB gesteuert, der über Kommunikationskanäle mit den technischen Anlagen verbunden ist, welche die SRL bereitstellen. Bei Abrufung der SRL wählt der ÜNB denjenigen Anbieter aus, der nach den Maximen der betrieblichen Netzsicherheit, der Versorgungssicherheit und der Kostenminimierung, am besten dazu geeignet ist.[41] Die Aktivierung der SRL beginnt ca. 30 Sekunden nach dem Auftreten einer Frequenzabweichung in der für das Ungleichgewicht verantwortlichen Regelzone. Sie hat zum Ziel, die Netzfrequenz auf den Sollwert zurückzuführen und den durch die PRL entstandenen außerfahrplanmäßigen Energieaustausch zwischen den Regelzonen rückgängig zu machen.[42] Dazu muss der sogenannte Area Control Error

[30] Consentec GmbH (2014, S. 19–20)
[31] Rebours et al. (2007b, S. 364)
[32] Rebours et al. (2007b, S. 364–365)
[33] Boje et al. (2014b, S. 7296–7297)
[34] Consentec GmbH (2014, S. 9–10)
[35] Verband der Netzbetreiber – VDN – e.V. beim VDEW (2003, S. 8)
[36] Rebours et al. (2007a, S. 353)
[37] ENTSO-E - European Network of Transmission System Operators for Electricity (2009, S. P1-4-P1-7)
[38] Rebours et al. (2007a, S. 353)
[39] Rebours et al. (2007b, S. 364)
[40] Rebours et al. (2007a, S. 353)
[41] Verband der Netzbetreiber – VDN – e.V. beim VDEW (2007, S. 52–53)
[42] Boje et al. (2014b, S. 7296–7297)

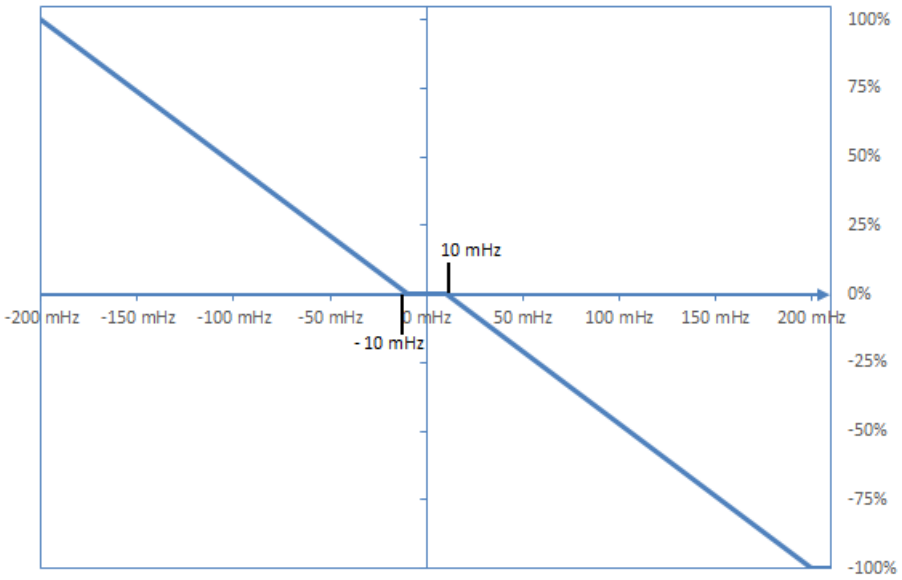

Abbildung 1: Aktivierung der PRL in Abhängigkeit von der Frequenzabweichung (Quelle: eigene Abbildung; vgl. Deutsche ÜNB (2014a))

(ACE) auf 0 ausgeglichen werden. Dieser besteht zum einen aus dem Saldo zwischen tatsächlichen Leistungsexport und fahrplanmäßigen Leistungsexport und zum Anderen aus dem Leistungsbeitrag der regelzoneneigenen PRL zur Frequenzstabilisierung. [43] Der ACE gibt folglich an, welche Leistung von der SRL zu erbringen ist, damit zum einen der tatsächliche Leistungsaustausch zwischen den Regelzonen wieder dem Fahrplan entspricht, wodurch faktisch die PRL der anderen Regelzonen abgelöst wird und zum anderen auch die von der Regelzone selbst erbrachte PRL heruntergefahren werden kann.

In der Vergangenheit (Jahre 2004 - 2005) wies Deutschland einen im europäischen Vergleich hohen RI-Wert für die SRL auf. Dies liegt an der Aufteilung Deutschlands in vier Regelzonen. Anders als die PRL, wird die SRL nicht solidarisch geleistet, sondern muss von jeder Regelzone eigenverantwortlich vorgehalten werden, was in gesamtdeutscher Sicht zu Überkapazitäten geführt hat.[44] Es ist jedoch anzunehmen, dass der deutsche RI-Wert für die SRL mit Einführung des gesamtdeutschen Netzregelverbundes 2010, der im folgenden Abschnitt samt seiner Wirkung beschrieben wird, gesunken ist.

2.5. Netzregelverbund

Zum 31.05.2010 ordnete die Bundesnetzagentur die Einführung eines Netzregelverbundes zwischen den vier deutschen Übertragungsnetzbetreibern und ihren Regelzonen an. Von nun an wurden Leistungsungleichgewichte zwischen den Regelzonen nicht mehr sofort verrechnet, sondern nur der Gesamtsaldo, der dann durch Regelenergie ausgeglichen wird. So lässt sich die für Regelenergie vorzuhaltende Leistung reduzieren und so Einsparungen von 16 Mio. Euro monatlich erzielen. Darin sind auch Kosteneinsparungen durch Wettbewerb zwischen den Anbietern von Regelleistung enthalten, der durch die, mit der Einführung des Netzregelverbundes einhergehende Bildung eines Gesamtmarktes für Regelenergie, verstärkt wurde.[45] Flinkerbusch und Heuterkes sind noch einen Schritt weiter gegangen und haben simuliert, welche Kosteneinsparung durch eine gänzliche Abschaffung der vier Regelzonen zugunsten einer einzigen deutschlandweiten Regelzone zu erzielen wäre. Insgesamt ergeben sich laut der Simulation Kosteneinsparungen von rund 163 Mio. Euro, was einer Reduktion um 17% entspricht. 160 Mio. Euro Einsparungen stammen dabei alleine aus der SRL, deren Aufwand sich, hauptsächlich durch die Saldierung der Abrufung zwischen den vereinigten Regelzonen, um 45% reduziert hat.[46] Die vier deutschen Regelzonen existieren zwar auch heute noch, dafür wurde der Netzregelverbund auf europäischen Nachbarländer ausgedehnt. Sobald das innerdeutsche Optimierungspotential ausgeschöpft ist, findet mittlerweile eine Saldierung von Leistungsungleichgewichten mit Regelzonen in Dänemark, den Niederlanden, der Schweiz, Tschechien und Österreich statt.[47,48]

2.6. Minutenreserveleistung

Die MRL stellt eine manuelle Änderung in der Kraftwerkseinsatzplanung dar, um PRL und SRL aus ihrem Dienst entlassen zu können, sodass die entsprechenden Kapazitäten für eine nächste Abrufung bereitstehen. Außerdem hat die

[43] Rebours et al. (2007a, S. 354)
[44] Rebours et al. (2007b, S. 364)
[45] Bundesnetzagentur (2010, S. 1–2)
[46] Flinkerbusch und Heuterkes (2010, S. 4717–4718)
[47] Aufzählung nach Beitrittsreihenfolge
[48] Deutsche ÜNB (2014a, S. 1–2)

MRL die Aufgabe die Nennfrequenz wiederherzustellen und, falls die SRL dazu nicht in der Lage war, die außerplanmäßig getätigen Energieaustausche auszugleichen.[49] Wie auch bei der Abrufung der SRL wählt der ÜNB den nach den Maximen der betrieblichen Netzsicherheit, der Versorgungssicherheit und der Kostenminimierung am besten geeigneten Anbieter aus.[50] Die MRL wird ca. 15 Minuten nach Auftreten der Frequenzabweichung aktiviert und kann in der Regel nach 45 - 75 Minuten von günstigeren Erzeugungseinheiten abgelöst werden.[51] Bei der Aktivierung erfolgt eine Anpassung im Fahrplan der entsprechenden Erzeugungseinheiten. Sie ist deren Betreiber mindestens 7 1/2 Minuten vor Beginn des nächsten Viertelstundenintervalls mitzuteilen, zu dem der geänderte Fahrplan in Kraft treten soll. Der Anbieter hat daraufhin 15 Minuten Zeit die angeforderte Minutenreserveleistung zu erbringen.[52] Die MRL ist im Gegensatz zu PRL und SRL ein vorwärtsgewandter Kontrollmechanismus, da er im Vorfeld vom ÜNB angefordert wird. Dieser trägt dadurch allerdings auch das Kostenrisiko einer Fehlaktivierung. PRL und SRL werden hingegen durch rückkoppelnde Regelkreise gesteuert und reagieren daher nur träge auf die Unausgewogenheit von Erzeugung und Verbrauch und können daher nur mit einer zeitlichen Verzögerung deren Gleichgewicht wiederherstellen.[53]

2.7. Fokussierung auf Primärregelleistung

Während PRL die Netzfrequenz nach einer Abweichung vom Sollwert nur stabilisieren soll, hat die SRL die Aufgabe, die außerplanmäßigen Leistungsflüsse zwischen den Regelzonen auszugleichen und den Sollwert der Netzfrequenz wiederherzustellen. Beiden ist dabei jedoch gemeinsam, dass sie diese Dienste nur für eine begrenzte Zeit erbringen. Die MRL ist dagegen auf einen längerfristigen Einsatz ausgelegt. Batteriespeicher, mit ihrer begrenzten Speicherfähigkeit, sind daher für die Erbringung von MRL nicht geeignet.[54] Roberts und McDowall stellen fest, dass vor allem kleinere Speicher, die ihre Leistung für Sekunden bis einige Minuten vorhalten, wirtschaftlich erfolgreich sind, da sie anders als konventionelle Erzeugungseinheiten sofort auf Ungereimtheiten im Netz reagieren können, aber trotzdem nicht als Konkurrenz für diese Anlagen gedacht sind, sondern vielmehr als eine Ergänzung.[55] Diese Erfolgsfaktoren der Energiespeicher lassen sich auf dem Gebiet der Regelleistung am besten bei der PRL mit ihren vergleichsweise kurzen Abrufedauern umsetzen. Auch Oudalov et al. kamen zu dem Ergebnis, dass die Bereitstellung von PRL die profitträchtigste Anwendung für einen Batteriespeicher ist: In ihrer Untersuchung verglichen die Autoren die Anschaffungs- und Betriebskosten eines Batteriespeichersystems mit dem zu erzielenden Umsatz durch die Bereitstellung verschiedener Systemdienstleistungen. Schlussendlich stellten sie fest, dass die Bereitstellung von PRL den meisten Wert für den Betreiber eines Batteriespeichers hat.[56] Daher soll im weiteren Verlauf dieser Bachelorarbeit von den verschiedenen Regelleistungsarten nur die Primärregelleistung für die Bereitstellung durch Batteriespeicher weiter betrachtet werden. Im folgenden Abschnitt werden die in Deutschland geltenden Anforderungen an die Anbieter von PRL genauer erläutert.

2.8. Anforderungen an Erbringer von PRL in Deutschland

2.8.1. Regelungen des TransmissionCode 2007

Um PRL vermarkten zu können, muss ein Anbieter neben des Abschlusses eines Rahmenvertrages über die Bereitstellung der Regelleistung auch Anforderungen technischer und betrieblicher Art erfüllen, die im Online-Appendix D1 des TransmissionCodes 2007 festgehalten sind: Dazu zählt an erster Stelle die Pirmärregelfähigkeit. Für Erzeugungseinheiten mit einer Nennleistung über 100 MW ist sie sogar eine Grundvoraussetzung für den Netzanschluss. Unter Pirmärregelfähigkeit versteht man unter anderem, dass mindestens +/- 2% der Nennleistung einer Anlage als Bandbreite der PRL vom ÜNB abgerufen werden können. Die Statik der Abrufung muss dabei je nach Anweisung des ÜNB einstellbar sein. Zur Primärregelfähigkeit gehört zudem die zeitliche Verfügbarkeit der vereinbarten Leistung. So muss ab einer gemessenen Frequenzabweichung von +/- 200 mHz innerhalb von 30 Sekunden die volle PRL gleichmäßig aktiviert und für mindestens 15 Minuten verfügbar sein. Die zulässige Aktivierungsdauer verkürzt sich entsprechend bei kleineren Frequenzabweichungen.[57] Die volle PRL muss nicht nur 15 Minuten lang abgegeben, sondern, beispielsweise beim Wegfallen einer großen Last, auch aufgenommen werden können. So ergibt sich für Anbieter von PRL ein Mindestleistungsvermögen von 30 Minuten Abrufung bei 100% der angebotenen Leistung. [58] Die Anbieter müssen, wenn Sie von ihrem ÜNB dazu aufgefordert werden, die ordnungsgemäße Erbringung der PRL beweisen können. Dazu übermittelt der Betreiber dem ÜNB Daten über seine Istleistung und den Frequenzverlauf der letzten zwei Wochen. [59] Zur Minimierung des Ausfallrisikos der PRL schreibt die ENTSO-E für die Beschaffung von Regelenergie eine homogene Verteilung der Anlagen vor. Von den 3 GW, die insgesamt im europäischen Verbundsystem zu beschaffen sind, dürfen maximal 3% von einer Erzeugungseinheit bzw. 6% von einem Netzknoten stammen. Falls ein ÜNB, die von ihm zu beschaffende PRL nicht durch Markttransaktionen aufbringen kann, ist er gegen Zahlung eines finanziellen Ausgleichs dazu berechtigt, in Betrieb befindliche Erzeugungsanlagen zur Bereitstellung von PRL zu verpflichten. [60]

[49] Rebours et al. (2007a, S. 351–352)
[50] Verband der Netzbetreiber – VDN – e.V. beim VDEW (2007, S. 52–53)
[51] Boje et al. (2014b, S. 7297)
[52] Verband der Netzbetreiber – VDN – e.V. beim VDEW (2007, S. 53)
[53] Weissbach und Welfonder (2009, S. 30)
[54] Borsche et al. (2013, S. 1)
[55] Roberts und McDowall (2005, S. 25)
[56] Oudalov et al. (2006, S. 2208, 2211)
[57] Verband der Netzbetreiber – VDN – e.V. beim VDEW (2007, S. 27–28)
[58] Borsche et al. (2013, S. 2)
[59] Verband der Netzbetreiber – VDN – e.V. beim VDEW (2003, S. 10)
[60] Verband der Netzbetreiber – VDN – e.V. beim VDEW (2007, S. 51)

2.8.2. Vorgaben der Bundesnetzagentur bezüglich der Ausschreibung von Primärregelleistung

Im April 2011 veröffentlichte die Bundesnetzagentur neue Regeln für die Ausschreibung von Regelenergie, die bedeutende Änderungen unter anderem hinsichtlich der Mindesangebotsgröße und der Auschreibungszeiträume brachte. Der Beschluss BK6-10-097 trat am 27.06.2011 in Kraft, ersetzte dabei die Regelungen des BK6-06-65 und ist bis heute gültig. Zudem überschreiben die Regelungen der Bundesnetzagentur, als oberste Kontrollinstanz, die Vorgaben der ÜNBs, sofern diese gegensätzlich sind. Im Folgenden werden einige der Regeln erläutert und mit den zuvor geltenden Regeln in Zusammenhang gebracht.[61]

Der Auschreibungsszeitraum für die PRL wird von einem Monat auf eine Woche reduziert. Diese Maßnahme ermöglicht mehr Flexibilität bei der Angebotserstellung und baut so, insbesondere für kleine Anbieter, eine Hürde für den Markteintritt ab. Auch die Kosten für die angebotene PRL sollen durch die Verkürzung des Ausschreibungszeitraums sinken. Zum einen verringern sich die von den Anbietern eingepreisten Marktrisiken. Zum anderen können die Anbieter gezielter die Anlagen für die Bereitstellung von PRL einsetzen, die in der jeweiligen Woche am kostengünstigsten sind. Durch diese Optimierungsmöglichkeit wird der Wettbewerb unter den Anbietern um das günstigste Angebot verstärkt. Darüber hinaus ist auch bei einem einwöchigen Ausschreibungszeitraum für PRL die Kompatibilität mit anderen Stromprodukten gesichert, da an der Strombörse EEX bereits im Jahr 2010 Wochenprodukte eingeführt wurden.[62] Die Bundesnetzagentur verfolgt damit eine Strategie weiter, die sie bereits im Jahr 2007 mit dem Inkrafttreten des Beschlusses BK6-06-065 eingesetzt hatte. Damals wurde der Ausschreibungszeitraum von sechs Monaten auf einen Monat herabgesetzt, ebenfalls mit dem Ziel mehr Wettbewerb im Primärregelmarkt zu erzeugen.[63] Die Bundesnetzagentur erwägt sogar eine weitere Verkürzung des Ausschreibungszeitraums auf nur einen Tag, so dass der Primärregelmarkt auch dargebotsabhängigen Erzeugern, wie WEA, offen steht. Eine Voraussetzung dafür ist jedoch, dass die Bereitstellung von PRL von solchen Anbietern zunächst Marktreife erreicht. Denn nur dann ist zu erwarten, dass diese potentiellen neuen Anbieter das Minderangebot der etablierten Akteure, resultierend aus dem zusätzlichen organisatorischen und abwicklungstechnischen Mehraufwand einer täglichen Ausschreibung für die Anbieter, ausgleichen können. [64] Die Ausschreibung selbst findet immer am Dienstag der Vorwoche statt. Während die Anbieter der PRL, zur Minimierung ihrer Risiken, eine sehr kurzfristige Ausschreibung bevorzugen würden, muss den ÜNBs genug Zeit für den Ausgleich einer möglichen Bedarfsunterdeckung bei der Versteigerung eingeräumt werden, weswegen Donnerstag und Freitag für die Ausschreibung nicht in Frage kommen. Da man eine zeitliche Kollision mit der Ausschreibung der SRL am Mittwoch vermeiden wollte, ergab sich der Dienstag als eine Kompromisslösung. Die Angebote können bis 15 Uhr abgegeben werden, wodurch eine zeitliche Überschneidung mit der Ausschreibung der Minutenreserve und dem Handel am Spotmarkt, die jeweils vormittags stattfinden, vermieden werden kann.[65]

Ein weiterer essentieller Bestandteil des Beschlusses BK6-10-097 ist die Herabsetzung der Mindestangebotsgröße von vormals +/- 5 MW auf nur noch +/- 1 MW. So erhalten mehr kleinere Anbieter einen Zugang zum Primärregelmarkt und Anbieter, die auf Grund der bisherigen Anforderungen, nur als ein Teil eines Anlagenpools am Primärregelmarkt teilnehmen konnten, haben nun die Möglichkeit selbstständig am Markt aufzutreten.[66] Die Bundesnetzagentur geht dabei auf eine Forderung ein, die unter anderem die Evonik Degussa GmbH vorgebracht hatte. Diese kritisierte, dass die bestehenden Regelungen zu sehr an den Gegebenheiten von Kraftwerken ausgerichtet sind. So seien aber auch Lithium-Ionen-Batterien hervorragend zur Bereitstellung von PRL geeignet. Zur Erprobung solch neuer Technologien sind in der Regel zunächst Testanlagen mit geringerer Kapazität nötig, die eine geringere Mindestangebotsgröße bedingen.[67] Auch in diesem Fall wird eine Maßnahme des Beschlusses BK6-06-065 weiter verstärkt. Damals wurden die von den ÜNBs geforderte Mindestangebotsgröße von +/- 10 MW von der Bundesnetzagentur auf +/- 5 MW herabgesetzt. Auch die Mindestanlagenleistung von 2 MW wurde abgeschafft, während die Möglichkeit zur Poolung von Anlagen verschiedener Anbieter eingeführt wurde. Genau wie durch die beschriebenen Maßnahmen aus dem Jahr 2011 sollte auch damals mehr kleineren Anbietern die Teilnahme an den Ausschreibungen für PRL ermöglicht werden, um so die Konkurrenz auf dem Primärregelmarkt zu erhöhen.[68]

Im Beschluss BK6-10-097 wurden auch einige Regeln des BK6-06-065 unverändert übernommen. So gibt es weiterhin keine Unterteilung der PRL in Zeitscheiben, wie beispielsweise eine Haupt- und eine Nebenzeit mit unterschiedlichen Bedarfsmengen. Denn der Wechsel zwischen den Zeitscheiben hätte die Umsortierung der regelleistungserbringenden Anlagen und damit eine eingeschränkte Verfügbarkeit der PRL zu diesem Zeitpunkt zur Folge. Ebenfalls aus Gründen der Systemsicherheit erfolgt bei der PRL keine Differenzierung nach der Regelrichtung.[69] Dies hat für die Anbieter von PRL zur Folge, dass sie ihre angebotene Leistung je nach Verhalten der Netzfrequenz ans Netz abgeben oder von selbigen aufnehmen können müssen. Die PRL wird daher auch bei ihrer Ausschreibung als symmetrisches Produkt beschafft. Bei SRL und MRL dagegen kann unabhängig voneinander positive und negative Regelleistung angeboten werden.[70] Auf wei-

[61] Bundesnetzagentur (2011, S. 1–4)
[62] Bundesnetzagentur (2011, S. 22–25)
[63] Bundesnetzagentur (2007, S. 8–9)
[64] Bundesnetzagentur (2011, S. 28–29)
[65] Bundesnetzagentur (2011, S. 32–33)
[66] Bundesnetzagentur (2011, S. 39)
[67] Bundesnetzagentur (2011, S. 14)
[68] Bundesnetzagentur (2007, S. 18–19)
[69] Bundesnetzagentur (2011, S. 35–36)
[70] Consentec GmbH (2014, S. 21–22)

tere Details des Ausschreibungsverfahrens, wie beispielsweise die Vergütung, wird in Kapitel 5.2, im Rahmen der Wirtschaftschaftlichkeitsberechnung, eingegangen.

2.9. Primärregelleistung im Ausland

Dass man bei der Vergütung von PRL durchaus andere Wege gehen kann, wird bei einem Blick ins Ausland deutlich. Die Ausgestaltung von PRL ist in den verschiedenen Ländern sehr unterschiedlich. Selbst in Dänemark, das sich mit Deutschland in einem Netzregelverbund befindet (vgl. Kapitel 2.1.3) und dessen westliches Teilnetz sich synchron zum kontinentaleuropäischen Netz verhält, gibt es große Unterschiede im Vergleich zu Deutschland: Die Versteigerung der PRL findet am Vortag statt. Der kommende Tag wird dabei in 4-Stunden-Blöcken vermarktet, wobei das höchste noch angenommene Angebot den Preis für alle anderen Angebote setzt. Für positive und negative PRL kann getrennt geboten werden, sowohl hinsichtlich der Menge der angebotenen Leistung als auch dem Preis. Die positive PRL hat dabei in der Regel einen höheren Preis als die negative. Grundsätzlich werden außerdem in den Nachtstunden höhere Preise erzielt als am Tag. Die angebotene Leistung muss genau wie in Deutschland (vgl. Kapitel 2.7.2) mindestens 15 Minuten vorgehalten werden.[71] Insgesamt bietet der dänische Markt den Anbietern von PRL somit mehr Flexibilität bei der Angebotserstellung.

Hingegen ist in dem ebenfalls zum Zuständigkeitsbereich von ENTSO- E gehörenden Spanien sowie Teilen der USA die Erbringung von PRL verpflichtend für bestimmte Erzeugerklassen und wird nicht vergütet. Dieses System kann trotzdem als fair für die Erzeuger betrachtet werden, da alle Erzeugungseinheiten einer bestimmten Klasse den gleichen absoluten oder relativen Beitrag zur PRL erbringen müssen.[72]

3. Batterietechnik

Bei Energiespeichern unterscheidet man grundsätzlich zwischen physikalischen Speichern wie Schwungrädern und Kondensatoren und den chemischen Energiespeichern, zu denen neben den Brennstoffzellen auch Batterien gehören. Bei den Batterien gibt es eine Unterteilung zwischen den aufladbaren Sekundärbatterien und den nicht-wiederaufladbaren primären Systemen. Die Batterien setzen sich aus einer bestimmten Anzahl von Batteriezellen zusammen. Bei mindestens einer Sekundärzelle spricht man von einem Akkumulator. Ab zwei Sekundärzellen kann man auch von einer Sekundärbatterie sprechen. Im alltäglichen Sprachgebrauch werden die Begriffe Akkumulator und Batterie häufig synonym verwendet. Die einzelnen galvanischen Zellen wandeln chemische in elektrische Energie um und bestehen mindestens aus einer positiven und einer negativen Elektrode, einem Separator, der beide elektrisch voneinander isoliert und so einen Kurzschluss verhindert, sowie einem Elektrolyt, der die Elektroden umgibt.[73] Beim Entladen findet an der Anode genannten Minuselektrode eine Oxidation und an der Kathode genannten Pluselektrode eine Reduktion statt. Für diese Redoxreaktionen ist das sogenannte Aktivmaterial der Elektroden zuständig. Sie finden an dessen Grenzschicht zum Elektrolyten statt und haben meist eine Volumenänderung der Elektroden zur Folge. Beim Laden kehrt sich der chemische Prozess um und der Strom fließt in umgekehrter Richtung. Dabei müssten die Bezeichnungen der Elektroden eigentlich vertauscht werden. Um Verwirrungen vorzubeugen wird die Zuordnung von Anode und Kathode vom Entladefall auch für die Ladung der Batterie beibehalten. Ein weiterer Bestandteil der Elektroden ist der Ableiter, der mit dem Aktivmaterial meist nicht identisch ist, sondern vielmehr als Trägermedium für selbiges wirkt. Daneben hat er die Aufgabe, die bei den Reaktionen benötigten bzw. freiwerdenden Elektronen zuzuführen bzw. abzuleiten. Der Ableiter ist meist metallisch und sollte sowohl eine chemische Widerstandsfähigkeit gegenüber dem Elektrolyten aufweisen, als auch mechanisch stabil sein. [74]

Eine wichtige Kenngröße für Batterien ist ihre Energiedichte. Sie gibt an wie viel Energie in Wattstunden (Wh) eine Batterie entweder im Bezug auf ihr Volumen oder ihrer Masse speichern kann. In letzterem Fall spricht man von der spezifischen Energiedichte, die in Wh/kg angegeben wird. Um das Speichervermögen von Batterien auszudrücken wird häufig auch die in Amperestunden (Ah) angegebene Kapazität verwendet. Daneben ist die Leistungsdichte eine weitere wichtige Kenngröße für Batterien. Je größer sie ist, desto besser ist die Batterie in der Lage große Energiemengen in kurzer Zeit abzugeben. [75] Je nach ihrer Optimierung sind Batterien entweder auf hohe Energiemengen oder eine hohe Leistung ausgelegt, jeweils zum Nachteil des anderen Parameters. So können hohe Energiedichten durch den Einsatz von leichten Aktivmaterialien, die den Austausch möglichst vieler Elektronen ermöglichen, und durch eine möglichst kleine Auslegung des Stromableiters, hinsichtlich Masse und Volumen, erreicht werden. Hohe Leistungsdichten werden hingegen durch möglichst große Elektrodenoberflächen, bei gleichzeitig möglichst kleinen Abstand zwischen ihnen, und durch einen möglichst kleinen ohmschen Innenwiderstand der Bauteile erreicht. Zusätzlich wirkt sich die Wahl von Elektrodenmaterialien, die eine möglichst große Zellspannung ergeben, positiv für beide Optimierungsrichtungen aus.[76]

Ein weiterer wichtiger Parameter für die Eigenschaften von Batterien ist die Temperatur bei der sie betrieben werden. So führt ein Temperaturanstieg zu einer Erhöhung der Leitfähigkeit des Elektrolyten. Gleichzeitig sinkt die Leitfähigkeit des Ableiters. Da dessen Auswirkungen auf den ohmschen Gesamtwiderstand allerdings gering sind, führt ein Temperaturanstieg zu einer Erhöhung der Leitfähigkeit der Batterie und letztlich auch zu einer Erhöhung der tatsächlich nutzbaren Kapazität. Bei einer Temperaturerhöhung laufen die Lade- und Entladereaktionen schneller ab, wobei die

[71] Swierczynski et al. (2013, S. 488, 491)
[72] Rebours et al. (2007b, S. 359–361)
[73] Birke und Schiemann (2013, S. 18, 26–32, 79)
[74] Retzbach (2008, S. 31–32)
[75] Retzbach (2008, S. 47–50)
[76] Jossen und Weydanz (2006, S. 29–30, 81)

Faustformel gilt, dass sich die Reaktionsgeschwindigkeit bei einem Temperaturanstieg von 10°C verdoppelt. Das Gleiche gilt jedoch leider auch für die unerwünschten Nebenreaktionen, die die Selbstentladung der Batterie verstärken und ihre Alterung beschleunigen. Vor allem ab Temperaturen über 50°C erreichen die Batterien ihr Lebensende viel schneller.[77]

Allgemein gilt das Ende der Lebensdauer einer Batterie als erreicht, wenn diese nur noch 80% ihrer Nennkapazität aufweist. Man differenziert dabei zwischen der kalendarischen und der nutzungsbezogenen Lebensdauer. Bei Ersterer besteht eine große Abhängigkeit vom Ausmaß der Nebenreaktionen. Der Hauptgrund für eine nutzungsbedingte Alterung sind Volumenänderungen an den aktiven Elektroden. Vor allem von ihnen hängt ab, nach welcher Anzahl Zyklen die Batterie das Ende ihrer Lebensdauer erreicht.[78] Im Folgenden wird der Aufbau, das Funktionsprinzip und Eigenschaften von Bleisäurebatterien und Lithium-Ionen-Batterien erläutert. Im anschließenden Kapitel 4 werden dann Batteriespeichersysteme mit diesen beiden Technologien auf ihre Eignung für die Bereitstellung von PRL untersucht.

3.1. Bleisäurebatterien

Bis auf den Elektrolyt aus verdünnter Schwefelsäure und dem Separator, der in der Regel aus Polyethylen oder Polypropylen besteht, sind die meisten Bauteile der Bleisäurebatterie aus Bleiverbindungen, was das Recycling am Ende der Lebensdauer erleichtert. Die gitterförmigen Stromableiter sind aus reinem Blei oder einer Bleilegierung gefertigt. Das Aktivmaterial wird an der negativen Elektrode von Blei (Pb) und an der Kathode von Bleidioxid (PbO_2) gebildet. Bei Bleisäurebatterien sind theoretisch spezifische Energiedichten von bis zu 161 Wh/kg möglich. Die für eine hohe Energiedichte nötigen dünnen Stromableiter neigen jedoch stärker zu Gitterkorrosion, wodurch die Lebensdauer verkürzt wird. So sind in der Praxis Energiedichten von 25-45 Wh/kg üblich.[79]

$$Pb + PbO_2 + 2H_2SO_4 \underset{Ladung}{\overset{Entladung}{\rightleftharpoons}} 2PbSO_4 + 2H_2O \qquad (1)$$

Gesamtreaktion für die Ladung und Entladung einer Bleisäurebatterie

Beim Entladevorgang (vgl. Gleichung 1) werden beide Aktivmaterialien in eine Kristallstruktur aus Bleisulfat ($PbSO_4$) umgewandelt. Die für die Reaktion nötigen Sulfationen stammen aus dem schwefelsäurehaltigen Elektrolyt (H_2SO_4) und dienen gleichzeitig als Ladungsträger. Nahe der vollständigen Entladung steigt daher, auf Grund der kaum noch vorhanden Leitfähigkeit, der Innenwiderstand stark an, was zu starken Verzögerungen beim anschließenden Wiederaufladen führt.[80] Eine vollständige Entladung ist daher schon aus operativen Gründen nicht empfehlenswert. Neben der beabsichtigten Entladung tritt bei Bleisäurebatterien auch eine Selbstentladung auf, die 3% der Nennkapazität pro Monat beträgt. Dies gilt für eine Zelltemperatur von 20°C. Da die Selbstentladung von der Intensität der ungewollten Nebenreaktionen abhängt, verdoppelt sie sich bei einem Temperaturanstieg von 10°C. (vgl. Kapitel 3)

Bei den Bleisäurebatterien mit geschlossener Bauart tritt beispielsweise eine Gasung genannte Nebenreaktion auf, bei der, unter Verbrauch von Wasser, Sauerstoff und Wasserstoff gebildet wird.[81] Diese haben eine mit einem Stopfen verschlossene Öffnung, durch die das bei der Gasung entstehende Gemisch entweichen und gegebenenfalls Wasser nachgefüllt werden kann. Die Bleisäurebatterien der sog. verschlossenen Bauart bedürfen keiner solchen Öffnung: Durch die Rekombination des sich bildenden Wasserstoffs und Sauerstoffs kann ein Wasserverbrauch verhindert werden. Diese Systeme können daher als wartungsfrei bezeichnet werden. Da ein Transport des Gases, auf Grund seiner geringen Löslichkeit, nicht im flüssigen Elektrolyt erfolgen kann, muss dieser in einem Porensystem gebunden werden, um den für die Rekombination nötigen Austausch des Gases zwischen den Elektroden zu ermöglichen. Bei den Blei-Gel-Batterien wird der Elektrolyt in einem Gel aus Siliziumdioxid gebunden, bei den Blei-Vlies-Batterien hingegen in einem Vlies. Letztere Bauart weist nicht nur eine im Vergleich mit anderen Bleisäurebatterien hohe Energiedichte auf, sondern, auf Grund eines hohen Automatisierungsgrades bei der Herstellung, auch niedrige Produktionskosten. Darüber hinaus ist auch der im Vergleich zu Blei-Gel-Batterien niedrigere Innenwiderstand vorteilhaft.[82]

Zu den Vorteilen der Bleisäurebatterien zählen, neben dem günstigen Preis und der hohen Zellenspannung, auch der große Erfahrungsschatz im Betrieb dieses Batterietyps. Die Schwachpunkte dieser Batterietechnologie sind dagegen unter anderen die mit bis zu 40 Wh/kg eher klein ausfallende Energiedichte und die hohe Empfindlichkeit gegenüber tiefen Entladungen und hohen Zyklenzahlen. [83] Zum Wohle ihrer Lebensdauer sollten Bleisäurebatterien keinen hohen Temperaturen ausgesetzt sein und auch nicht längere Zeit im entladenen Zustand gehalten werden. Vielmehr ist die Chemie dieses Batterietyps im vollgeladenen Zustand am stabilsten. Daher sollten Bleisäurebatterien, die sich unter normalen Betriebsbedingungen in einem halbgeladenen Zustand befinden, in mehrwöchigen Abständen einmal voll geladen werden.[84] Bleisäurebatterien werden zur Zeit vor allem zum Sicherstellen einer unterbrechungsfreien Stromversorgung eingesetzt. Obwohl die gefluteten (geschlossene Bauweise) Bleisäurebatterien am zuverlässigsten sind, haben die verschlossenen Bautypen bei den Notstromversorgungssystemen, unter anderem wegen ihrer geringen Unterhaltskosten und ihrer kompakten Bauweise, die Marktführung inne.[85] Oudalov

[77] Jossen und Weydanz (2006, S. 27)
[78] Retzbach (2008, S. 55–56)
[79] Jossen und Weydanz (2006, S. 33–36, 46–47)
[80] Birke und Schiemann (2013, S. 70–72)
[81] Jossen und Weydanz (2006, S. 52)
[82] Birke und Schiemann (2013, S. 82, 88–90)
[83] Retzbach (2008, S. 69–70)
[84] Retzbach (2008, S. 86–88)
[85] Joseph und Shahidehpour (2006, S. 4)

et al. halten für große Batteriespeichersysteme ebenfalls den Einsatz der verschlossenen Bauweise, wegen ihrer hohen Effizienz und den geringen Anschaffungskosten, für vorteilhaft, auch wenn ihre Lebensdauer vergleichsweise gering ist.[86]

3.2. Lithium-Ionen-Batterien

Die Lithium-Ionen-Zellen beruhen auf einem anderem chemischen Grundprinzip als beispielsweise die eben beschriebenen Bleisäurebatterien, bei denen eine sogenannte Rekonstitution stattfindet: Beim Laden bzw. Entladen wird das Aktivmaterial in einen anderen Stoff umgewandelt. Bei Lithium-Ionen Batterien entstehen im Betrieb keine neuen Phasen, vielmehr werden gasförmige Lithium-Ionen nur in einem Wirtsgitter der Elektroden eingelagert. Das Funktionsprinzip ist in Abbildung 2 veranschaulicht: Beim Laden geben die in der Kathode eingelagerten Lithiumatome ein Elektron ab und wandern als positiv geladene Lithium-Ionen durch den lithium-ionen-haltigen Elektrolyt zur negativen Elektrode, werden dort unter Aufnahme eines Elektrons neutralisiert und im Wirtsgitter der Anode eingelagert. Diese Einlagerung der Lithiumatome wird als Interkalation bezeichnet. Beim Entladen findet dieser Prozess wegen der Stromumkehr in der entgegengesetzten Richtung statt. Da dieses hin und her wandern der Lithium-Ionen an die Bewegung eines Schaukelstuhls erinnert wurden die Lithium-Ionen-Batterien früher häufig als "Rocking Chair"-Batterien bezeichnet.[87]

Eine Besonderheit der Lithium-Ionen-Batterien ist die Vielzahl an Materialkombinationen, die für die Elektroden verwendet werden können. Dieser Umstand macht die Lithium-Ionen-Batterien sehr vielseitig einsetzbar, da sie durch die Wahl der Elektrodenmaterialien an den jeweiligen Einsatzzweck angepasst werden können. Grundsätzliche Stärken all dieser Systeme sind ein hoher energetischer Wirkungsgrad von bis zu 95% und eine hohe Energiedichte, die durch die hohen Zellspannungen von bis zu 4,2 Volt ermöglicht wird. Die Ableiter der Elektroden sind an der Kathode meist aus Aluminium und auf der negativen Seite in der Regel aus Kupfer gefertigt. Als Material für den Separator wird, wie bei den Bleisäurebatterien, häufig Polyethylen und Polypropylen verwendet.[88]

Wichtig für den Durchbruch der Lithium-Ionen-Technologie war das Finden eines Elektrolyten, der über weite Temperaturbereiche stabil ist, den Ableiter der positiven Elektrode nicht angreift und dazu noch eine hohe Leitfähigkeit aufweist. Diese Eigenschaften erfüllt eine Kombination aus Lithiumhexafluorophosphat als Leitsalz und einem organischen Lösungsmittelgemisch, das unter anderem Ethylencarbonat und Propylencarbonat enthält, bisher am Besten.[89] Ebenfalls entscheidend für den heutigen Erfolg war der Einsatz von Elektrodenmaterialien, die begünstigend für die Interkalation sind. Eine Schlüsseleigenschaft ist dabei die sogenannte Stöchiometriebreite, die bezogen auf die Lithium-Ionen-Zellen aussagt, wie viele Lithiumatome reversibel in den Elektroden eingelagert werden können, ohne dass es dadurch zu größeren Änderungen in der Struktur des Wirtsgitters kommt, die eine verringerte Zyklenlebensdauer zur Folge hätten. Eine hohe Stöchiometriebreite ermöglicht somit die hohe Energiedichte der Lithium-Ionen-Batterien, da diese mit der Zahl der Lithiumatome, die im Elektrodengitter eingelagert werden können, steigt.[90]

Im Folgenden werden einige gängige Aktivmaterialien für Anoden und Kathoden von Lithium-Ionen-Batterien vorgestellt.

Auf Seiten der Anode ist Graphit als Aktivmaterial dominierend. Es zeichnet sich vor allem durch sein sechseckiges, sehr flaches Kohlenstoffgitter aus, dessen Schichtabstand sich durch die Interkalation von Lithium-Atomen nur um 9% vergrößert. Da das Laden bzw. Entladen somit als hochgradig reversibler Prozess abläuft sind Lebensdauern von über 1.000 Zyklen möglich. Vorteilhaft ist, neben der ausreichenden Leitfähigkeit, außerdem die spezifische Kapazität von bis zu 380 mAh/g, vor allem im Vergleich zu ca. 200 mAh/g des amorphen Kohlenstoffs, der zur Markteinführung der Lithium-Ionen-Batterie häufig als Anodenmaterial verwendet wurde.[91] Ein neuartiges Elektrodenmaterial für die Anode stellt das sogenannte Lithium-Titanat ($Li_4Ti_5O_{12}$) dar. Durch die, im Vergleich zu Graphit, niedrigere Zellspannung ergibt sich eine auf 150 mAh/g reduzierte Speicherkapazität. Die Stärken des Lithium-Titanat liegen in seiner Beständigkeit. Da bei der Interkalation keinerlei Volumenänderung auftritt und es außerdem sehr robust gegenüber falscher Handhabung ist, können bis zum Lebensende mehrere tausend Zyklen durchlaufen werden.[92] Während auf Grund der geringen Zellspannung und Energiedichte ein Einsatz von Lithium-Titanat im mobilen Bereich unwahrscheinlich erscheint, ist es besonders für Anwendungen geeignet, die ein hohes Maß an Stabilität und Zuverlässigkeit erfordern.[93]

Auf der Seite der Kathoden ist aus historischen und praktischen Gründen das Lithium-Cobaltoxid ($LiCoO_2$) das meist verwendete Aktivmaterial. Es zeichnet sich durch eine hohe spezifische Kapazität von 150 mAh/g und eine hohe Potentialdifferenz von 4,2 Volt gegenüber Graphit aus. Ein Überschreiten dieses Spannungswertes führt zu einer Materialzersetzung in Verbindung mit einer stark exothermen Reaktion und stellt somit ein Sicherheitsrisiko dar. Als weitestgehend Eigensicher gilt dagegen das Lithium-Mangan-Oxid ($LiMn_2O_4$; LMO), das dafür jedoch nur eine spezifische Kapazität von 120 mAh/g erreicht. Ein neuer Ansatz ist die Verwendung von Mischkathoden, bestehend aus den Metalloxiden Nickel, Mangan und Kobalt ($Li(Ni_xCo_yMn_z)O_2$; NMC). Bei dieser Variante kann über den Anteil der einzelnen Metalloxide die Ausprägung der jeweiligen Eigenschaften ein-

[86] Oudalov et al. (2006, S. 2206)
[87] Jossen und Weydanz (2006, S. 104–105)
[88] Birke und Schiemann (2013, S. 193–198)
[89] Birke und Schiemann (2013, S. 203–204)
[90] Birke und Schiemann (2013, S. 204–207)
[91] Retzbach (2008, S. 132–133)
[92] Retzbach (2008, S. 133–134)
[93] Jossen und Weydanz (2006, S. 112–113)

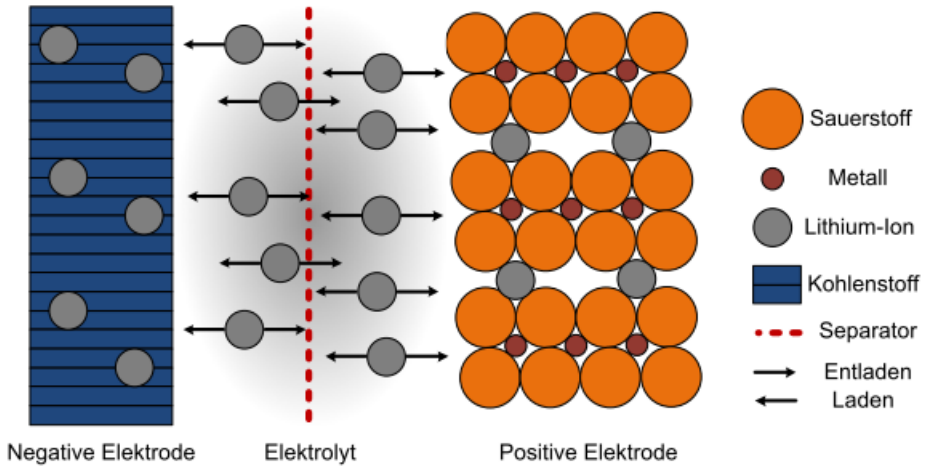

Abbildung 2: Funktionsprinzip der Lithium-Ionen Batterien (Quelle: Fuchs et al. (2012))

gestellt werden.[94] Die spezifische Kapazität liegt so zwischen 130-160mAh/g. [95] Ein weiteres, sehr vielversprechendes Kathodenmaterial ist Lithium-Eisen-Phophat ($LiFePO_4$; LFP). Es zeichnet sich neben seiner hohen spezifischen Kapazität von 160 mAh/g vor allem durch seine überragende Langzeitstabilität aus: Wichtiger als die Tolerierung begrenzter Überspannungen ist dabei die geringe interkalationsbedingte Volumenänderung von nur 7%. So sind Lebensdauern von mehreren tausend Zyklen möglich. Ein weiterer Vorteil des LFP ist die hohe Verfügbarkeit und der günstige Preis des benötigten Rohmaterials.[96] Nachteilig an LFP ist dagegen seine geringe Leitfähigkeit. Diese kann jedoch durch einen Überzug mit Kohlenstoff erhöht werden.[97]

Bei der Selbstentladung sind die Lithium-Ionen Batterien, mit einer Rate von 2-3% der Kapazität pro Monat bei Zimmertemperatur, allen anderen Batterietypen überlegen. Neben hohen Temperaturen und fortgeschrittenen Zellalter verstärken auch die hohen Ladespannungen nahe der Volladung die Selbstentladung, da sie die Intensität der Nebenreaktionen verstärken.[98] Zum Wohle der Lebensdauer sollten daher Vollladungen vermieden werden. Ideal für die Lebensdauer von Lithium-Ionen-Batterien ist ein Ladezustand zwischen 30-70%.[99] Diese Anforderung ist mit den Einsatzbedingungen bei der Bereitstellung von PRL sehr gut vereinbar: Die Auflage jederzeit positive und negative Regelleistung anbieten zu können (vgl. Kapitel 2.8.2), bedingt für den anbietenden Batteriespeicher einen mittleren Ladezustand von 50%.

3.3. Nickel-Cadmium- und Nickel-Metalhydrid-Batterien

Eine weitere bekannte Batterietechnologie sind die Nickel-Cadmium-Batterien (NiCd). Doch obwohl sie gegenüber Bleisäurebatterien einige Vorteile aufweisen, wie eine hohe Energiedichte und ein gutes Entladeverhalten bei tiefen Temperaturen, sind NiCd-Batterien wegen des giftigen Schwermetalls Cadmium nahezu vom Markt verschwunden. Auch die Nickel-Metalhydrid-Batterien (NiMH) werden trotz einer hohen Energie- und Leistungsdichte zunehmend aus dem Markt verdrängt: Der Preis von Lithium-Ionen-Batterien ist in den letzten Jahren stark gesunken und gleicht sich immer mehr dem Preis für NiMH-Batterien an. Dabei ist diese neue Technologie den NiMH-Batterien in einigen Punkten überlegen: So können Lithium-Ionen-Batterien auf Grund ihrer hohen Spannungslage Energiedichten von bis zu 150 Wh/kg erzielen, während NiMH-Batterien nur rund 80 WH/kg erreichen. Auch der Bereich der nutzbaren Ladungszustände ist bei Lithium-Ionen-Batterien größer und die Zyklenlebensdauer höher als bei NiMH-Batterien.[100] NiCd- und NiMH-Batterien sollen daher zugunsten der Lithium-Ionen-Technologie in dieser Arbeit nicht weiter behandelt werden.

4. Batteriespeichersysteme für die Erbringung von Primärregelleistung

4.1. Batteriemanagementsystem

Normalerweise sind Batterien direkt mit der Einheit verbunden, in der die eingespeicherte Energie verbraucht wird. Batteriespeicher, die Netzdienstleistungen wie die PRL erbringen sollen, sind jedoch mit dem Wechselstromnetz verbunden. Die dazu nötigen großformatigen Batterieströnge müssen dazu durch ein Batteriemanagementsystem (BMS) ergänzt werden, das in der Regel aus einer Steuereinheit für die Batterieströnge, einem Temperaturregelungssystem und einem Umrichter besteht. Durch die Kombination der Batterie mit dem BMS entsteht ein modulares, zuverlässiges und effizientes Batteriespeichersystem (BSS).[101]

[94] Retzbach (2008, S. 134–135) So lässt sich die Überladestabilität des Manganoxids mit der hohen Kapazität des Cobaltoxids vereinen.
[95] Jossen und Weydanz (2006, S. 117)
[96] Retzbach (2008, S. 135–136)
[97] Scrosati und Garche (2010, S. 2424)
[98] Retzbach (2008, S. 148)
[99] Jossen und Weydanz (2006, S. 150–151)
[100] Birke und Schiemann (2013, S. 120, 153, 232–233)
[101] Mercier et al. (2009, S. 1470)

Bei der Entladung der Batterie hat das BMS die folgenden Aufgaben zu erfüllen: Während durch die Steuereinheit des BMS die Spannung und die Stromstärke für eine maximale Leistungsabgabe eingestellt werden kann, hat der Umrichter die Aufgabe den aus der Batterie ausgespeicherten Gleichstrom in Wechselstrom umzuwandeln und mit dem Wechselstrom des Übertragungsnetzes zu synchronisieren.[102] Das Temperaturregelungssystem wird mit wachsenden Stromstärken immer wichtiger, da die resultierende Erhitzung der Zellen schädlich für die Lebensdauer des BSS ist. Neben einer generellen Erwärmung soll vor allem eine ungleichmäßige Erwärmung verhindert werden.[103]

Für die Anschaffungskosten des BMS gelten andere Grundsätze als für die Batteriestränge. Während sich die nötige Auslegungsgröße, und damit auch die Kosten, bei einem Batteriespeicher primär nach der geforderten Energiekapazität richtet, hängt die Auslegung des Umrichters des BMS nur von der gewünschten Nennleistung ab.[104] Insgesamt können die Kosten für das BMS mehr als 25% der Gesamtkosten eines BSS ausmachen. Derzeit gibt es aber eine rasante technologische Entwicklung auf diesem Gebiet, die vor allem durch den Vormarsch der EE getrieben ist, bei denen ähnliche Managementsysteme für den Netzanschluss erforderlich sind. So ist zu erwarten, dass die BMS zukünftig noch verlässlicher arbeiten und gleichzeitig günstiger in der Anschaffung sind.[105]

4.2. Eigenschaften von Batteriespeichersystemen

Bevor auf die Eignung der BSS für die Erbringung von PRL eingegangen wird, sollen an dieser Stelle zunächst anhand der Eigenschaften der BSS die Unterschiede zu den thermischen Kraftwerken herausgearbeitet werden, die bisher den überwiegenden Teil der Regelleistung erbringen (vgl. Kapitel 2).

Ein entscheidender Unterschied liegt in der Natur der BSS als Speichereinheit. Während die Leistungsabgabe bei konventionellen Kraftwerken unabhängig von früheren Betriebszuständen erfolgt, ist sie bei BSS von der zu einem früheren Zeitpunkt eingespeicherten Energie abhängig. Die Möglichkeit zur Einspeicherung, also zur Aufnahme von Energie aus dem Netz, fehlt wiederum den konventionellen Kraftwerken. Während den BSS durch ihre Lade- und Entladevorgänge somit zwei Betriebsmodi zur Verfügung stehen, gibt es bei thermischen Kraftwerken nur den Zustand der Leistungsabgabe.[106] Ein weiterer Unterschied ergibt sich aus der sehr kurzen Reaktionszeit der BSS: Ihnen ist es möglich innerhalb von Millisekunden ihre Nennleistung abzugeben oder aufzunehmen. Bei konventionellen Erzeugungseinheiten kommt es dagegen zu Verzögerungen von 15 - 30 Sekunden.[107]

Weitere positive Eigenschaften von BSS sind zum einen ihre Modularität, die eine einfache Skalierung ermöglicht, und zum anderen ihre Unabhängigkeit von geographischen Gegebenheiten, die sie überall einsetzbar macht. Nachteilig sind hingegen, trotz sinkender Preise, weiterhin die hohen Anschaffungskosten von BSS im Vergleich zu ihrer Energiekapazität. So werden sie vor allem für Anwendungen rentabel sein, die hohe Nennleistungen im Vergleich zur Speicherkapazität erfordern.[108] Die Bereitstellung von PRL ist, wegen der kurzen Abrufungszeiten von maximal einer halben Stunde (vgl Kap. 2.8.1), dementsprechend eine potentielle Anwendung für BSS. Unabhängig davon, ob BSS nach heutigen Maßstäben für die Erbringung von PRL geeignet sind, könnten sie in Zukunft eine Notwendigkeit für die Sicherstellung der Systemstabilität sein. Dies wird im folgenden Abschnitt genauer erläutert.

4.3. Neue Herausforderungen als Chance für Batteriespeichersysteme

Eine wichtige Eigenschaft der konventionellen Kraftwerke für die Systemstabilität ist ihre Rotationsträgheit. Sie entsteht durch die in den rotierenden Massen der Generatoren gespeicherte kinetische Energie. Diese wird bei einem Absinken der Netzfrequenz abgegeben. Im umgekehrten Fall kann auch zusätzliche kinetische Energie vom Netz aufgenommen werden. Die Rotationsträgheit der konventionellen Kraftwerke trägt somit zur Stabilisierung des Netzes bei. Da die WEA und PV-Anlagen über Wechselrichter mit dem Stromnetz verbunden sind und daher keine Rotationsträgheit besitzen, führt der Ausbau der EE grundsätzlich zu einem Rückgang an Rotationsfähigkeit und zudem zu einer Fluktuation der Rotationsträgheit, da die Einspeiseleistung von WEA und PV-Anlagen stark über die Zeit schwankt. Lag die Rotationsträgheit früher noch konstant bei 6 Sekunden, sinkt sie heutzutage zeitweise auf 3 - 4 Sekunden ab. Die Frequenz ändert sich also schneller und erfordert daher auch eine schnellere Frequenzregulierung.[109] Bisher wird Anlagen, die PRL erbringen 30 Sekunden Zeit zum Hochfahren eingeräumt. Der Verlust an Rotationsträgheit kann jedoch Anlagen erforderlich machen, die die Regelleistung schneller erbringen können. Auch aus ökonomischen und ökologischen Gründen sollten die konventionellen Kraftwerke zunehmend durch andere Anbieter von Regelleistung ersetzt werden. Um die PRL zu erbringen müssen konventionelle Kraftwerke zumindest im Teillastbetrieb gehalten werden. Durch den Ausbau der EE laufen diese Kraftwerke daher zu manchen Zeiten nur, um PRL bereitstellen zu können und verursachen dadurch unnötige Kosten und einen CO_2-Ausstoß.[110] BSS dagegen können dank ihrer zwei Betriebsmodi die Regelleistung aus einer Ruhepostion heraus erbringen und reagieren dabei innerhalb von Millisekunden (vgl. Kapitel 4.2). Während BSS also für zukünftige Herausforderungen des Energienetzes bestens geeignet sind, soll im folgenden Abschnitt dargestellt werden,

[102] Joseph und Shahidehpour (2006, S. 4)
[103] Jossen und Weydanz (2006, S. 245)
[104] Anderson und Carr (1993, S. 476)
[105] Divya und Østergaard (2009, S. 514)
[106] Lo und Anderson (1999, S. 825)
[107] Oudalov et al. (2006, S. 2208)
[108] Koller et al. (2013, S. 1)
[109] Boje et al. (2014b, S. 7290–7293)
[110] Boje et al. (2014a, S. 4038)

ob sie sich auch nach heutigen Gesichtspunkten schon für die Erbringung von PRL eignen.

4.4. Eignung für die Bereitstellung von Primärregelleistung

Batteriespeicher für die Bereitstellung von Primärregelleistung zu nutzen ist keinesfalls eine neue Idee. So wurde bereits 1987 in Westberlin ein auf Bleisäure basierendes BSS, mit einer Speicherkapazität von 14,4 MWh und einer Leistung von 8,5 MW für die Frequenzregelung errichtet. Kunisch et al. kamen damals schon zu der Einschätzung, dass BSS zwar für andere Systemdienstleistungen nicht so gut geeignet, aber, sowohl aus praktischer als auch ökonomischer Sicht, vielversprechend für die Bereitstellung von PRL im Berliner Inselsystem[111] sind.[112] Heutzutage ist das gesamte deutsche Energiesystem Teil des kontinentaleuropäischen Verbundnetzes. Daher soll in dieser Arbeit die Eignung und Wirtschaftlichkeit von BSS für die Bereitstellung von PRL für ein solches Umfeld untersucht werden. Ein Einsatz in Inselsystemen soll dagegen nicht weiter betrachtet werden.

Ein Einsatz von BSS für die Bereitstellung von PRL macht laut Borsche et al. nur Sinn, wenn dadurch die Systemstabilität nicht gefährdet wird. So muss trotz der begrenzten Speicherkapazität von BSS sichergestellt sein, dass sie das Referenzereignis (vgl. Kapitel 2.2) jederzeit, also auch wenn bereits Leistung zum Laden bzw. Entladen verwendet wird, meistern können: Für einen Zeitraum von 15 Minuten ist die volle Abrufleistung zu erbringen (vgl. Kapitel 2.8.2). Direkt nach einer PRL-Abrufung kann bei den BSS zur Aufrechterhaltung der Einsatzbereitschaft eine Ladung bzw. Entladung nötig sein. Um die Systemstabilität sicher zu stellen, müssen die Anbieter von SRL daher genügend Leistungskapazität aufweisen, um neben ihrer regulären Abrufung auch diese zusätzliche Leistungsanforderung zu stemmen.[113] Laut Boje et al. entsteht durch den Einsatz von BSS zur Bereitstellung von PRL jedoch kein zusätzlicher Beschaffungsbedarf an SRL. In der von ihnen durchgeführten Ausfallrechnung konnten die BSS außerdem in jedem Szenario ihre Eignung beweisen. Die BSS seien daher für die Erbringung von PRL genauso verlässlich wie konventionelle Kraftwerke.[114]

Nach den Regelungen der Bundesnetzagentur muss die PRL immer in gleicher Höhe für beide Regelrichtungen angeboten werden (vgl. Kapitel 2.8.3). Damit BSS diese Vorgabe erfüllen können, also jederzeit die volle angebotene Leistung abgeben oder aufnehmen können, ergibt sich für diese ein mittlerer Ladezustand von 50%. Batterietechnologien, die für die Erbringung von PRL eingesetzt werden, sollten daher bei dieser Betriebsweise effizient arbeiten können. Laut Swierczynski et al. sind die Lithium-Ionen-Batterien sehr gut für die Erbringung von PRL geeignet. Denn gerade Ladestände um die 50% sind ideal für die Lebensdauer dieses Batterietyps. Darüber hinaus können sie in kurzen Zeiträumen von bis zu 15 Minuten sehr hohe Leistungen abgeben.[115] Auch Koller et al. sehen für ein BSS zur Erbringung von Regelleistung vor allem die Lithium-Ionen-Technologie als geeignet an, da sie eine hohe Effizienz haben und die geforderte Leistung innerhalb von Sekunden erbringen können. So schafft in der Praxis ein in Zürich errichtetes BSS mit einer Nennleistung von 1 MW den Wechsel von der maximalen negativen zur maximalen positiven Abrufung in weniger als einer Sekunde. In den ersten zwei Jahren des Betriebs des Speichers konnten auch, abgesehen von einer technisch unvermeidbaren bei der ersten Ladung, keine weiteren Kapazitätseinbußen festgestellt werden.[116]

Anders als bei Lithium-Ionen-Batterien ist für Bleisäurebatterien ein mittlerer Ladungszustand (State of Charge; SoC) von 50% nicht ideal für die Lebensdauer (vgl. Kapitel 3.1). Eine Vergrößerung der Zyklenlebensdauer lässt sich bei der verschlossenen Bauform aber durch einen Überzug der Anode mit Kohlenstoff erreichen. Laut Doughty et al. basiert ein Großteil der Energiespeicher, die für die Erbringung von Regelleistung eingesetzt werden, auf Bleisäurebatterien. [117] Allerdings werden laut Fuchs et al., wegen des großen Aufsehens um die Lithium-Ionen-Technologie, Bleisäurebatterien als Basis für zukünftige BSS nicht hinreichend berücksichtigt. Dabei sprechen die recht günstigen Betriebskosten und vor allem die vergleichsweise geringen Investitionskosten für einen Einsatz der Bleisäuretechnologie zur Bereitstellung von PRL.[118] In dieser Arbeit soll daher neben einem fiktiven BSS mit Lithium-Ionen-Technologie auch eine entsprechende Anlage mit Bleisäurebatterien auf ihre Wirtschaftlichkeit bei der Erbringung von PRL überprüft werden.

Die Eignung von BSS zeigt sich auch darin, dass bereits Batteriespeicher in Deutschland den Präqualifizierungsprozess erfolgreich durchlaufen haben und PRL bereistellen. Beispielsweise die Anlage der Wemag AG in Schwerin mit einer Speicherkapazität von 5 MWh und einer Nennleistung von 5 MW. Der Erfolg dieses Projektes lässt sich jedoch nicht verallgemeinern, da es durch die Bundesregierung mit Fördergeldern in Höhe von 20% der Investitionskosten von 6,5 Millionen Euro unterstützt wurde.[119] Vor kurzem sind zudem, durch neue Regelungen der ÜNB, die im folgenden Abschnitt erläutert werden, die Anforderungen an BSS, die PRL erbringen wollen, gestiegen. In dieser Arbeit soll daher die Wirtschaftlichkeit unter normalen Marktbedingungen und unter Einhaltung der neuen Regelungen untersucht werden.

4.5. Zusatzanforderungen für Batteriespeicher

Am 26.08.2015 haben die ÜNB auf ihrer gemeinsamen Internetplattform www.regelleistung.net neue Anforderungen für BSS veröffentlicht, die am Markt für PRL teilnehmen wollen. Nach der Neuregelung müssen die Betreiber von BSS einen Vollabruf für 30 Minuten, in postiver als auch negativer

[111] Westberlin war auf Grund der politischen Umstände zur damaligen Zeit ein Inselsystem
[112] Kunisch et al. (1986, S. 41–42, 46)
[113] Borsche et al. (2013, S. 3)
[114] Boje et al. (2014a, S. 4043)
[115] Swierczynski et al. (2013, S. 488)
[116] Koller et al. (2015, S. 128–130)
[117] Doughty et al. (2010, S. 49)
[118] Fuchs et al. (2012, S. 43)
[119] Kempkens (2014a)

Richtung, sicherstellen können. Die deutschen ÜNB richten sich dabei nach den zukünftigen Anforderungen des "Network Code on Load-Frequency-Control and Reserves", der in der Europäischen Union eingeführt werden soll.[120]

In der Neuregelung wurden außerdem für die BSS zulässige Ladezustandsbereiche, je nach deren Auslegungsverhältnis von Speicherkapazität und angebotener Leistung, festgelegt. Dieser Arbeitsbereich des Ladezustandes muss im sogenannten Normalbetrieb, bei dem die Frequenzabweichungen grundsätzlich kleiner als 50 mHz sind, jederzeit eingehalten werden. Dies muss vorab durch Simulationen bewiesen werden. Der Arbeitsbereich darf nur bei Auftreten der in Tabelle 1 beschriebenen Fälle verlassen werden. Spätestens 2 Stunden nach der Rückkehr in den Normalbetrieb muss sich der Ladezustand wieder im Arbeitsbereich befinden. Die obere und untere Grenze des Arbeitsbereiches errechnet sich aus der sog. nutzbaren Speicherkapazität und der angebotenen Leistung, wie in den Formeln 2 und 3 dargestellt. Die sog. nutzbare Speicherkapazität muss dabei im Rahmen eines von den ÜNB festgelegten Testverfahrens ermittelt werden. Für ein BSS mit einer nutzbaren Speicherkapazität von 1,25 MWh und einer angebotenen Leistung von 1 MW ist somit ein Ladezustand zwischen 60% und 40% zulässig.[121]

$$C_{oG} = \frac{E_{nutzbar} - 0,5h \cdot P_{angeboten}}{E_{nutzbar}} \quad (2)$$

$$C_{uG} = \frac{0,5h \cdot P_{angeboten}}{E_{nutzbar}} \quad (3)$$

Obere und untere Grenze des Arbeitsbereichs

Für zulässige Ladestrategien zur Einhaltung des Arbeitsbereiches verweist das Dokument auf die "Eckpunkte und Freiheitsgrade bei der Erbringung von Primärregelleistung". Die darin enthaltenen Vorgehensweisen werden in Kapitel 5.4 erläutert. Damit auch während der Lade- und Entladevorgänge, die zur Sicherstellung des Ladezustands im Arbeitsbereich nötig sind, eine volle Abrufung der PRL möglich ist, muss die Leistungselektronik um das 1,25 fache größer ausgelegt sein als die angebotene Leistung.[122]

Auch wenn ein BSS alle der aufgeführten Anforderungen erfüllt, ist eine Präqualifikation nur auf Widerruf möglich, da es sich bei den Regelungen um vorläufige handelt, die jederzeit durch neue Anpassungen ersetzt werden können. Zudem sollen zukünftige Präqualifizierungskriterien auch für bestehende Anlagen gelten.[123] Diese abschließende Bemerkung stößt bei Tobias Struck, dem Leiter des Bereiches "Energiespeicher und Projekte" der Wemag AG auf Kritik. Sie bedeute einen großen Unsicherheitsfaktor für Betreiber von Batteriespeichern, vor allem auch für zukünftige Speicherprojekte. Die erhöhten Anforderungen für Batteriespeicher hält er für unbegründet, da sie weder von der Wissenschaft für nötig erachtet, noch im Ausland so praktiziert werden.[124] Falls sich dieser Verdacht der Unnötigkeit erhärten sollte, stellen die neuen Regelungen der ÜNB einen Verstoß gegen § 23 EnWG dar, der vorschreibt, dass die Regelungen zum Ausgleich des Energieversorgungsnetzes sachlich gerechtfertigt sein müssen. Falls sie nicht sachlich gerechtfertigt sind, würden die Neuregelungen zudem die Betreiber von Batteriespeichern gegenüber anderen Anbietern von PRL diskriminieren und daher gegen eine weitere Vorgabe des § 23 EnWG verstoßen.[125]

5. Wirtschaftlichkeitsberechnung für ein fiktives Batteriespeichersytems

In Kapitel 4 wurden die wichtigsten Eigenschaften von BSS und ihre Eignung für die Bereitstellung von PRL dargestellt, sowie die damit einhergehenden Zusatzanforderungen, erläutert. Nun soll für ein fiktives, auf Bleisäurebatterien bzw. auf der Lithium-Ionen-Technologie basierendes BSS jeweils die Wirtschaftlichkeit überprüft und abschließend vergleichend dargestellt werden. Viele der bisher realisierten Batteriespeichersysteme zur Bereitstellung von PRL sind nicht rentabel zu betreiben, da sie überdimensioniert sind. Sie wurden nämlich nicht nach ökonomischen Gesichtspunkten ausgelegt, sondern sollten vor allem die technische Machbarkeit solcher Systeme verdeutlichen.[126] So hat auch ein Energiespeicher, der zur Zeit in Aachen errichtet wird, neben der Erbringung von PRL, vor allem die Forschung an weiteren Anwendungen für BSS und die Erprobung verschiedener Batterietechnologien zum Ziel.[127] Das fiktive BSS in dieser Abschlussarbeit soll dagegen ganz nach wirtschaftlichen Gesichtspunkten ausgelegt werden, um herauszufinden, ob BSS unter Marktbedingungen profitabel PRL bereitstellen können.

Für einen wirtschaftlichen Betrieb eines BSS ist es entscheidend, die Speicherkapazität für eine gegebene bereitgestellte Leistung zu minimieren. Für einen optimalen Betrieb sollten daneben auch die Auslegung der Leistungskapazität, operative Kosten, wie die Beschaffung von Energie zum Nachladen der Batterie, und die Alterung des BSS möglichst klein gehalten werden.[128] Zur Profitmaximierung kann auch die Kombination des Batteriespeichers mit einer konventionellen Erzeugungseinheit beitragen, sodass die jeweiligen Stärken der Systeme vereint werden können. So würde bei der Erbringung von PRL die reaktionsschnelle Batterie nach kurzer Zeit abgelöst, so dass nur eine geringe Speicherkapazität erforderlich ist. Des weiteren ließe sich durch das gleichzeitige Anbieten mehrerer Dienstleistungen der erwirtschaftete Umsatz bei gleichbleibenden Kapitalkosten er-

[120] Deutsche ÜNB (2015, S. 3)
[121] Deutsche ÜNB (2015, S. 4, 7)
[122] Deutsche ÜNB (2015, S. 5)
[123] Deutsche ÜNB (2015, S. 9)
[124] vgl. Online-Appendix A2 - Wemag AG
[125] Deutscher Bundestag (2005)
[126] Mercier et al. (2009, S. 1469)
[127] vgl. Online-Appendix A3 - RWTH Aachen
[128] Borsche et al. (2013, S. 1)

Tabelle 1: Ausnahmen vom Normalbetrieb

	Frequenzabweichung	**Mindestdauer**
Ausnahme 1)	größer 200mHz	-
Ausnahme 2)	größer 100mHz	5 Minuten
Ausnahme 3)	größer 50mHz	15 Minuten

höhen.[129] Diese zwei Möglichkeiten zur Profitmaxmierung werden in dieser Arbeit allerdings nicht weiter verfolgt, da untersucht werden soll, ob ein eigenständiges BSS nur durch die Bereitstellung von PRL rentabel betrieben werden kann.

5.1. Investitions- und Betriebskosten

Die Minimierung der Speicherkapazität ist deshalb so wichtig, da von ihr für eine gegebene Nennleistung die Investitionskosten direkt abhängen. Die Speicherkapazität stellt somit den Hauptkostentreiber eines BSS dar.[130] Um der zentralen Rolle der Investitionskosten bei der Wirtschaftlichkeitsbetrachtung Rechnung zu tragen, wurden für sie aktuelle Informationen von Herstellern und Betreibern von BSS eingeholt. In dem dazu erstellten Fragebogen wurden auch weitere Informationen über die Bereitstellung von PRL durch Batteriespeicher abgefragt.[131] Beispielsweise welche Batterietechnologie aus Sicht des Unternehmens sich am besten dafür eignet und wie hoch die erwartete Lebensdauer eines solchen BSS ist. Der Fragebogen wurde den adäquaten Ansprechpartnern per Mail übermittelt und im Fall der Wemag AG im Rahmen eines Telefoninterviews durchgesprochen. Die aus den Antworten resultierenden relevanten Erkenntnissen zu den Investitionskosten eines BSS sind in Tabelle 2 zusammengefasst.

Bei der Betrachtung der Werte fällt auf, dass sich mit einer Verdopplung der Nennleistung und der Speicherkapazität auch die Investitionskosten nahezu verzweifachen. Es sind somit fast keine Skaleneffekte ersichtlich. Laut Thomas Idinger von der Siemens AG treten Skaleneffekte spürbar erst ab Nennleistungen größer 10 MW auf.[132] Gleichzeitig führt eine Verdopplung der angebotenen Leistung zu einer Verzweifachung des Ertrages aus der Bereitstellung von PRL (siehe Kapitel 5.2). Insgesamt würde sich somit nahezu die gleiche Rentabilität ergeben wie im einfachen Fall. Daher soll für die weitere Wirtschaftlichkeitsbetrachtung nur die Speicherkapazität variiert und eine Nennleistung von 1,25 MW für das fiktive BSS festgesetzt werden. Dieser Wert ergibt sich aus der Mindestangebotsgröße für die Erbringung von PRL von 1 MW (vgl. Kapitel 2.8.2) und einem Leistungspuffer von 25% der angebotenen Leistung, den die Regelungen der deutschen ÜNB bezüglich BSS, die PRL erbringen, vorsehen (vgl. Kapitel 4.5). Des Weiteren soll die Wirtschaftlichkeitsberechnung auf Seiten der Lithium-Ionen Technologie auf die Variante mit LFP als Kathodenmaterial fokussiert werden, da diese die günstigsten Investitionskosten der drei Lithium-Ionen-Varianten aufweist. Unter der Annahme der linearen Entwicklung der Investitionskosten ergibt sich abhängig vom Verhältnis der Speicherkapazität zur Nennleistung (Kapazitäts-Leistungs-Verhältnis; KLV) ein fester Wert für die spezifischen Investitionskosten in €/kWh. Die Veränderung dieser Kosten in Abhängigkeit vom KLV lässt sich in einer Formel darstellen (siehe Formel 4). Mit Hilfe der Formel und den resultierenden spezifischen Investitionskosten lassen sich so die absoluten Investitionskosten für verschiedene KLV, mit einer Nennleistung von 1,25 MW als Basis, darstellen (siehe Tabelle 3).

$$k_{Li} = -150 \cdot KLV + 1275 \qquad k_{Pb} = -250 \cdot KLV + 1250 \qquad \left[\frac{\text{EUR}}{\text{kWh}}\right] \tag{4}$$

Spezifische Kosten von Lithium-Ionen (LFP) und Blei-Gel BSS in Abhängigkeit vom KLV

Die Betriebskosten für die Lithium-Ionen-Technologie wurden mit Hilfe der Angabe eines Herstellers[133] von BSS ermittelt. Sie betragen demnach jährlich rund 2,5% der Investitionskosten.[134] Nach Auskunft der Exide Technologies GmbH wachsen die Betriebskosten linear mit der Auslegungsgröße und betragen pro MWh bzw. pro MW 8.000 Euro.[135] Die berechneten Werte für die jährlichen laufenden Kosten für die verschiedenen Technologien und Auslegungsgrößen sind ebenfalls der Tabelle 3 zu entnehmen. Es sei darauf hingewiesen, dass die aufgeführten Betriebskosten noch keine Aufwendungen für Fahrplangeschäfte zum Laden des BSS beinhalten. Erwartungsgemäß hat die Blei-Gel-Technologie, sowohl bei der Anschaffung als auch im laufenden Betrieb, Kostenvorteile gegenüber den Lithium-Ionen-Batteriesystemen (vgl. Kapitel 4.4). Für die Zukunft prognostiziert das Fraunhofer-Institut sinkende Investitionskosten bei gleichzeitig steigendem Wirkungsgrad und einer höheren zyklischen sowie kalendarischen Lebensdauer. So sollen bis zum Jahr 2050, im Vergleich zu den Investitionskosten im Jahr 2012, die Preise für Lithium-Ionen- und Bleisäurebatterien um rund 80% fallen. [136] Aus heutiger Sicht ist für ein BSS, das PRL bereitstellt, neben den Investitionskosten, vor allem der Erfolg bei den Ausschreibungen zur PRL für die Profitabilität entscheidend, wie im folgenden Abschnitt erläutert wird.

[129] Jin et al. (2011, S. 1–2)
[130] Oudalov et al. (2007, S. 1259)
[131] vgl. Online-Appendix A-1 Fragenkatalog
[132] vgl. Online-Appendix A5 - Siemens AG
[133] Das Unternehmen hat um eine anonyme Behandlung seiner Daten gebeten
[134] vgl. Online-Appendix A6 - Herstellerangabe
[135] vgl. Online-Appendix A4 - Exide Technologies GmbH
[136] Fraunhofer-Institut für Windenergie und Energiesystemtechnik et al. (2014, S. 20)

Tabelle 2: Investitionskosten laut Angaben von Herstellern und Betreibern

[1] Der Hersteller hat um eine Anonymisierung gebeten
[2] Die Kostenangabe der Wemag AG bezog sich auf 5MWh/5MW und wurde linear heruntergebrochen
Vgl. auch Online-Appendix A5 - Siemens AG, Online-Appendix A6 - Herstellerangabe, Online-Appendix A2 - Wemag, Online-Appendix A4 - Exide Technologies GmbH

Anbieter/Betreiber	Technologie	1MWh/1MW	2MWh/1MW	2MWh/2MW
Siemens AG	Lithium-Ionen (NMC)	1.500.000€	2.500.000€	2.900.000€
Herstellerangabe[1]	Lithium-Ionen (LFP)	1.125.000€	1.950.000€	2.250.000€
Wemag AG[2]	Lithium-Ionen (LMO)	1.300.000€		2.600.000€
Exide Technologies GmbH	Blei- Gel	1.000.000€	1.500.000€	2.000.000€

Tabelle 3: Errechnete spezifische Kosten, Investitionskosten sowie jährliche Betriebskosten

	Auslegung (MWh/MW):	**1,0/1,25**	**1,125/1,25**	**1,25/1,25**	**1,375/1,25**	**1,5/1,25**
	KLV:	**0,8**	**0,9**	**1**	**1,1**	**1,2**
Lithium -Ionen (LFP)	spez. Kosten (€/kWh)	1.155,00€	1.140,00€	1.125,00€	1.110,00€	1.095,00
	Investitionskosten:	1.155.000,00€	1.282.500,00€	1.406.250,00€	1.526.250,00€	1.642.500,00€
	jährliche Betriebskosten	28.875,00€	32.062,50€	35.156,25€	38.156,25€	41.062,50€
Blei Gel	spez. Kosten (€/kWh)	1.050,00€	1.025,00€	1.000,00€	975,00€	950,00
	Investitionskosten:	1.050.000,00€	1.153.125,00€	1.250.000,00€	1.340.625,00€	1.425.000,00€
	jährliche Betriebskosten	18.000,00€	19.000,00€	20.000,00€	21.000,00€	22.000,00€

5.2. Erträge aus den Ausschreibungen zur Primärregelleistung

Bevor ein Anbieter Erträge aus der Bereitstellung von PRL erzielen kann, muss sein Angebot zunächst bei der wöchentlichen Ausschreibung (vgl. Kapitel 2.8.2) angenommen werden. Diese erfolgt auf der gemeinsam von den ÜNB betriebenen Internetplattform www.regelleistung.net. Dort werden nicht nur die Angebote, die jeweils aus einem Angebotsumfang in MW und einem in €/MW angegebenen Leistungspreis bestehen, sondern auch die Ausschreibungsergebnisse veröffentlicht. Die Bezuschlagung bzw. Ablehnung eines Anbieters richtet sich nach der Höhe des gebotenen Leistungspreises. Beginnend mit dem günstigsten Angebot erfolgt die Annahme nach steigendem Leistungspreis, bis die für die primäre Frequenzregelung benötigte Leistung erreicht ist.[137] Da die PRL regelzonenübergreifend ausgeschrieben wird,[138] handelt es sich dabei um den von der ENTSO-E für Deutschland festgelegten Beitrag zur kontinentaleuropäischen PRL in Höhe von 568 MW (vgl. Kapitel 2.2). Während bei SRL und MRL auch die Abrufung der Regelleistung über einen im Angebot des Anbieters spezifizierten Arbeitspreis vergütet wird, erfolgt bei der PRL nur eine Vergütung in Höhe des angebotenen Leistungspreises (pay-as-bid). In den Jahren 2008 bis 2013 stagnierte der durchschnittliche mittlere Leistungspreis um die 20 Euro pro MW und Stunde.[139] Auf den Ausschreibungszeitraum von einer Woche hochgerechnet ergibt sich so ein durchschnittlicher Ertrag von 3.360 Euro aus der Bezuschlagung bei der Ausschreibung der PRL.

Um die Erträge des fiktiven BSS aus der Bereitstellung von PRL für die Wirtschaftlichkeitsuntersuchung zu bestimmen, wurden die Auschreibungsergebnisse des Jahres 2013 ausgewertet. Die entsprechende Datensätze können auf www.regelleistung.net abgerufen werden und enthalten für jeden Ausschreibungszeitraum die Leistungspreise und Größe der bezuschlagten Angebote. Insgesamt wurden 52 Gebotsrunden berücksichtigt, beginnend mit der Woche vom 31.12.2012 - 06.01.2013 und abschließend mit dem Ausschreibungszeitraum vom 23.12.2013 - 29.12.2013. Anhand von Auswertungen der Daten in Microsoft Excel wurde eine ertragsoptimierte Bietstrategie für das fiktive BSS ermittelt.[140] Neben einer konservativen Strategie, die darauf ausgelegt ist, möglichst in jeder Gebotsrunde angenommen zu werden, wurde auch eine aggressive Strategie in Betracht gezogen, die Nichtbezuschlagungen in manchen Wochen zu Gunsten eines höheren Leistungspreises in Kauf nimmt. Ideal für den Ertrag wäre das Anbieten des größten, noch angenommenen Leistungspreises der aktuellen Ausschreibungsperiode. Dadurch ergäbe sich ein Jahresumsatz von 174.778 Euro und ein durchschnittlicher Leistungspreis von 3.361,12 Euro. Da jedoch die hierfür zu Grunde liegende Annahme von vollkommener Information nicht realistisch ist, wurde diese Bietstrategie nicht weiter berücksichtigt. Auf der anderen Seite ist ein Mitbieten für einen Ausschreibungszeitraum nur sinnvoll, wenn der erzielte Leistungspreis über den wöchentlichen Betriebskosten des BSS liegt, sofern man diese als variabel annimmt. Um diesen Sachverhalt für alle Gebotsrunden sicherzustellen, wurde der minimale Leistungspreis

[137] Jansen (2014, S. 27)
[138] Bundesnetzagentur (2011, S. 37)
[139] Consentec GmbH (2014, S. 22–24)
[140] vgl. Online-Appendix A7 -Gebote

des Jahres 2013, in Höhe von 2.098 Euro, mit den erwarteten Betriebskosten aus Tabelle 3 verglichen. Kosten für die Beschaffung von Energie zum Ausgleich des Ladungszustandes werden dabei nicht berücksichtigt, da sie unter realen Umständen zum Zeitpunkt der Angebotsabgabe nicht prognostizierbar sind. Im Ergebnis zeigt sich, dass ein Mitbieten in allen Gebotsrunden sinnvoll ist, da selbst die Verwendung der maximalen wöchentlichen Betriebskosten aller Varianten (789,66 € /Woche) in einem Überschuss von 1.308,34 Euro resultiert. Es sei noch hinzugefügt, dass die verwendeten Betriebskosten sicherlich auch fixe Bestandteile enthalten, die unabhängig von einer Teilnahme bzw. Annahme bei der PRL-Ausschreibung anfallen.

Die untersuchten Bietstrategien für das fiktive BSS basieren alle auf statistischen Parametern, wie Mittelwerte und Quantile, der Leistungspreise der angenommenen Angebote der Vorwochen und wären somit auch in der Realität praktizierbar. Für eine weitere Differenzierung der Bietstrategien wurden die Werte der statistischen Kenngrößen um bestimmte Prozentwerte erhöht bzw. erniedrigt. Die Prozentwerte wurden dabei so gewählt, dass sich auf Basis jedes verwendeten statistischen Parameters konventionelle und aggressive Bietstrategien ergeben [141]. Einige der 29 untersuchten Bietstrategien sind, nach dem Ertrag im Jahr 2013 geordnet, in Tabelle 4 dargestellt.

Es fällt auf, dass sich in der Spitzengruppe ausschließlich konservative Bietstrategien mit einer Annahmerate von 100% befinden. Am besten schneidet, mit einem Jahresertrag von 150.327,84 Euro, der um 4% verminderte Wert des 3. Quantils der in der Vorwoche akzeptierten Leistungspreise ab und wird daher als Basis für die weitere Wirtschaftlichkeitsüberprüfung verwendet. Die Ertragsunterschiede zwischen den konservativen Bietstrategien sind eher als marginal zu bezeichnen. So wird auch durch Bieten mit dem minimalen Leistungspreis der Vorwoche ein Ertrag von 145.516,00 Euro erwirtschaftet. Entscheidend für den wirtschaftlichen Erfolg ist also, möglichst für alle Ausschreibungszeiträumen bezuschlagt zu werden. Die höheren Leistungspreise einer aggressiven Bietstrategie können den Verlust aus daraus resultierenden Ablehnungen in den Gebotsrunden nicht auffangen. Das liegt vor allem auch an der relativ geringen Preisspanne zwischen höchsten noch akzeptierten und niedrigsten Leistungspreis eines Ausschreibungszeitraumes. Der Median dieser Preisdifferenz im Jahr 2013 liegt bei nur 246,50 Euro, während der minimale Leistungspreis im gleichen Zeitraum 2.098 Euro betrug. Die Preisspanne der Leistungspreise bietet also nur sehr wenig Spielraum zum Ausgleich von nicht akzeptierten Geboten. Die Wemag AG verfolgt mit ihrem BSS in Schwerin ebenfalls eine konservative Bietstrategie, da die Erbringung von PRL der einzige Einsatzzweck und damit auch einzige Einnahmequelle des Speichers ist. Durch seine langjährige Erfahrung im Energiehandel versucht das Unternehmen die Annahme im wöchentlichen Ausschreibungsverfahren sicherzustellen und gleichzeitig den erzielten Leistungspreis zu optimieren.[142] Auch die für das fiktive BSS ausgewählte Bietstrategie wurde in allen Gebotsrunden des Jahres 2013 akzeptiert und muss somit für das ganze Jahr entsprechend den Frequenzabweichungen vom Sollwert PRL erbringen. Im folgenden Abschnitt soll das Ausmaß dieser Aktivierungen von PRL und die resultierenden Auswirkungen auf das BSS dargestellt werden.

5.3. Aktivierung der Primärregelleistung im Jahr 2013

Für den weiteren Verlauf der Wirtschaftlichkeitsberechnung des fiktiven BSS zur Bereitstellung von PRL ist die Aktivierung von selbiger im beispielhaften Jahr 2013 eine zentrale Einflussgröße: Zum einen hängt von ihr der Umfang der Ausgleichsenergie ab, die benötigt wird, um den Ladezustand der Batterie im erlaubten Arbeitsbereich zu halten und die Einsatzbereitschaft des BSS für weitere Aktivierungen der PRL sicherzustellen. Da durch die Beschaffung der Ausgleichsenergie Kosten entstehen, beeinflusst der Umfang der Abrufung der PRL auch die Profitabilität des BSS. Zum anderen hängt von der Aktivierung der PRL (PRL-Aktivierung) auch ein großer Teil der Lade- und Entladezyklen und deren Tiefe ab, während für die restlichen Zyklen das Lademanagement des BSS verantwortlich ist. In diesem Zusammenhang soll ein Vollzyklus eine vollständige Entladung mit anschließender Vollladung bezeichnen. Auch das Lademanagement wird zum Teil von der PRL-Aktiverung beeinflusst, da der Einsatz einer der in Kapitel 5.4 vorgestellten Ladestrategien nur in Phasen ohne Abrufung von PRL erfolgen kann.

Für die Ermittlung der Aktivierung der PRL, konnte wegen ihrer direkten Abhängigkeit von der Frequenzabweichung von ihrem Sollwert (vgl. Kapitel 2.3), ein Datensatz ausgewertet werden, der die Netzfrequenz im europäischen Verbundnetz für jede Sekunde des Jahres 2013 enthält. Um mit den von der Internetseite des ÜNB 50 Hertz stammenden Daten[143] sinnvoll weiterarbeiten zu können, ist eine Aggregierung nötig, wobei sich eine minutenscharfe Aufbereitung der Daten anbietet: So kann der Datensatz in einer Microsoft Excel-Datei verarbeitet werden, wobei gleichzeitig durch die resultierenden 525.600 Zeilen immer noch ein hinreichend großer Detaillierungsgrad erhalten bleibt. Noch vor der Aggregierung wird die Netzfrequenz in den entsprechenden Aktivierungsgrad umgewandelt, so dass in der minutenscharfen Darstellung auch solche Aktivierungen berücksichtigt werden, die aus Frequenzverläufen nahe um die Grenze des Totband (vgl. Kapitel 2.3) herum resultieren. Eine direkte Aggregierung der Netzfrequenz könnte in diesen Fällen einen Wert innerhalb des Totbands als Ergebnis haben, so dass für die entsprechende Minute, entgegen der Realität, keine Aktivierung verzeichnet würde. Eine unvermeidbare Beeinträchtigung der Datenintegrität bei der Aggregierung entsteht durch Umkehrungen der Aktivierungsrichtung innerhalb von einer Minute, die von hier an als Intra-Min-Umkehr bezeichnet wird, da in solchen Fällen die positive

[141] vgl. Online-Appendix A7 - Gebote

[142] vgl. Online-Appendix A2 - Wemag AG

[143] Abrufbar unter http://www.50hertz.com/de/Maerkte/Regelenergie/Regelenergie-Downloadbereich

Tabelle 4: Konservative und aggressive Bietstrategien im Vergleich

Rang	Bietstrategie	Annahmen (#)	Annahmen (%)	Jahresertrag	Ertrag/Annahme
1	3.Quartil-Vorwoche -4%	52	100%	150.327,84€	2.890,92€
2	Mittelwert-Vorwoche -2,5%	52	100%	150.219,96€	2.888,85€
3	Median-Vorwoche -2,75%	52	100%	148.951,02€	2.864,44€
4	3.Quartil-Vorwoche -5%	52	100%	148.761,93€	2.860,81€
5	Media-Vorwoche -3%	52	100%	148.568,11€	2.857,08€
11	Minimum-Vorwoche	52	100%	145.516,00€	2.798,38€
27	Media-Vorwoche +5%	25	48%	75.413,10€	3.016,52€
28	3.Quartil-Vorwoche +5%	20	38%	62.398,09€	3.119,90€
29	Maximum-Vorwoche	17	33%	52.559,00€	3.091,71€

mit der negativen Abrufung aufgerechnet wird und so eine ungewollte Glättung der PRL-Aktivierung entsteht. Ein solche Umkehr trat in 4,56% der Minuten des Jahres 2013 auf, eine Diskrepanz zwischen der maximalen negativen und positiven Abrufung von mehr als 10% Punkten allerdings nur in 0,7% aller Minuten.[144] Die Intra-Min-Umkehr stellt somit keine ernsthafte Gefahr für die Datengüte dar. Bei der Aggregierung wurden für jede Minute des Jahres 2013 der minimale und maximale Wert, der Mittelwert, sowie der Median des Abrufungsgrades erfasst. Die weiteren Untersuchungen sollen primär auf Basis der Medianwerte der Abrufung durchgeführt werden, da so Außreißerwerte weniger stark ins Gewicht fallen als bei Verwendung von Mittelwerten[145]. Das passt gut zu den in der Realität vorhandenen Trägheiten, wie beispielsweise im Übertragungsnetz (vgl. Kapitel 4.3).

Über den Abrufungsgrad und die vom fiktiven BSS angebotene Leistung von 1.000 kW lässt sich für jede Minute des Jahres 2013 die eingespeicherte bzw. entladene Energie darstellen. Insgesamt wurde 195,04 MWh positive und 161,39 MWh negative Regelenergie aktiviert. Beim Aufaddieren der Energiemengen, wird ein recht deutlicher Überhang der positiven PRL von 33,65 MWh ersichtlich, was 17,25% der gesamten erbrachten positiven PRL entspricht.[146] Aus dem Ausspeicherüberhang ergibt sich auch automatisch die Mindestmenge an Energie, die als Ausgleichsenergie in das BSS eingespeichert werden muss. Ein allein durch die Aktivierung der PRL bedingter Ausgleich des Ladezustands auf seinen Ausgangswert, ist nur zu Beginn des Jahres zu beobachten, da im restlichen Zeitraum die positive Abrufemenge der negativen vorauseilt. Dieser automatische Ausgleich trat bis einschließlich 27.01. insgesamt nur zehnmal im Jahr 2013 auf. Für die Bestimmung der Anzahl der Vollzyklen, die das BSS im Jahr durchlaufen hat, ist nicht der Saldo von positiver und negativer PRL von Bedeutung, sondern vielmehr die Einzelsummen der Beiden. Ein entscheidender Faktor ist dabei die Zyklentiefe, die in Prozent aussagt, wie stark eine Batterie im Vergleich zu ihrer gesamten Speicherkapazität entladen wurde. In der Literatur kursieren verschiedene Definitionen dieser Entladetiefe. In dieser Arbeit soll sich die Entladetiefe nur auf halbe Zyklen beziehen, beispielsweise die Tiefe einer Entladung bezogen auf die Speicherkapazität. Eine nachfolgende Ladung der Batterie stellt schon den nächsten Zyklus mit eigener Zyklentiefe dar.[147] Für die Ermittlung der Zyklentiefen der PRL-Aktivierungen ist es zunächst erforderlich den Betrag der prozentualen Ladung bzw. Entladung für die einzelnen Minuten zu ermitteln. Danach werden diejenigen Minuten, die jeweils Teil der identischen PRL-Abrufung sind, in einem Zyklus zusammengefasst und dabei die Einzelwerte der Lade- bzw. Entladetiefen zur Zyklentiefe aufaddiert. Dabei sollen PRL-Aktivierungen in die gleiche Richtung, die nur durch eine Ruhephase, aber keine entgegengerichtete Abrufung unterbrochen werden, in einem Zyklus zusammengefasst werden. Insgesamt ergeben sich so für das Jahr 2013 23.614 Zyklen, die, auf Basis einer Speicherkapazität des BSS von 1 MWh, jedoch nur eine durchschnittliche Zyklentiefe von 0,75% aufweisen. So haben 76,73% der PRL-Aktivierungen eine Zyklentiefe von weniger als 1% der Speicherkapazität. Aufaddiert sind die einzelnen Zyklen äquivalent zu 89,11 Vollzyklen, die jeweils aus einer vollständigen Entladung und anschließender Vollladung des Batteriespeichers bestehen. In 43,98% der Minuten des Jahres 2013 erfolgte keine PRL- Aktivierung.[148] Diese Ergebnisse decken sich mit jenen einer Untersuchung von Oudalov et al. aus dem Jahr 2007, bei der die PRL-Aktivierung im kontinentaleuropäischen Verbundnetz des Jahres 2005 ausgewertet wurde. Ebenfalls auf Basis eines KLV von 1, hatten 80% der Zyklen eine Tiefe von weniger als 0,3% der Speicherkapazität.[149] Die eben erläuterten Ergebnisse der Auswertung der PRL-Aktivierung im Jahr 2013 sind in Tabelle 5 zusammengefasst.

Die in diesem Abschnitt ermittelte äquivalente Vollzyklenzahl auf Grund der PRL-Aktivierung ist ein Einflussfaktor für die Lebensdauer des BSS. Bisher nicht berücksichtigt wurden Zyklen, die durch das Lademanagement des BSS hervorgerufen werden. Diese sollen im folgenden Abschnitt ermittelt werden, in dem die nach den Regelungen der ÜNB

[144] vgl. Online-Appendix A8 - PRL-Aktivierung
[145] Lippe (1993)
[146] vgl. Online-Appendix A8 - PRL-Aktivierung
[147] Koller et al. (2013, S. 2)
[148] vgl. Online-Appendix A8 - PRL-Aktivierung
[149] Oudalov et al. (2007, S. 1262)

Tabelle 5: Aktivierung der PRL im Jahr 2013 und resultierende Zyklisierung eines 1MWh/1MW BS

Alle Auslegungsvarianten			
Positive Abrufung (MWh)	195,04	Intra-Min-Umkehr (% der Minuten)	4,55%
Negative Abrufung (MWh)	-161,39	– 10% Umkehr	0,70%
Abrufungssaldo (MWh)	33,65	– 20% Umkehr	0,12%
– Anteil an pos. Abrufung	17,25%	Keine Abrufung (% der Minuten)	43,98%
Auslegungsvariante: 1MWh Speicherkapazität & 1 MW Angebotene Leistung			
Zyklenzahl	23614	Maximale Zyklentiefe (% der Speicherkap.)	34,68%
äquivalente Vollzyklenzahl	89,11	Mittlere Zyklentiefe (% der Speicherkap.)	0,75%
Maximale Zyklendauer (Minuten)	337	Median der Zyklentiefe (% der Speicherkap.)	0,26%
Mittlere Zyklendauer (Minuten)	12,47	Anteil Zyklen mit Zyklentiefe $\leq$ 1%	76,73%

erlaubten Verfahren zum Lademanagement vorgestellt und in Abhängigkeit verschiedener Auslegungsgrößen des BSS angewendet werden.

5.4. Lademanagement

Da sich die Abrufemengen der positiven und negativen Regelenergie nicht ausgleichen (vgl. Kapitel 5.3), ist zwangsläufig ein vom Betreiber gesteuertes Laden bzw. Entladen des Energiespeichers nötig. Außerdem sehen die Regelungen der ÜNB bezüglich BSS, die PRL erbringen, nur einen begrenzten Bereich vor, in dem der Ladezustand schwanken darf (vgl. Kapitel 4.5) und verweisen für zulässige Ladestrategien auf die "Eckpunkte und Freiheitsgrade bei der Erbringung von Primärregelleistung", die im April 2014 von den deutschen ÜNB veröffentlicht wurden. Diese sehen Fahrplangeschäfte vor, um den Ladezustand innerhalb des Arbeitsbereiches zu halten bzw. dorthin zurückzukehren. Betreiber von Batteriespeichern haben so die Möglichkeit sich durch bilaterale Geschäfte oder über den Börsenhandel Energiemengen zu beschaffen oder zu verkaufen. Diese Geschäfte sind den ÜNB mindestens 15 Minuten im Voraus mitzuteilen, damit die entsprechende physikalische Erbringung dieser Energiemengen gewährleistet werden kann. Die Geschäfte selbst folgen ebenfalls einem viertelstündlichen Intervall. Zur Unterstützung dieses Batterielademanagements darf im Totband PRL erbracht werden, allerdings nur wenn ein systemkonformes Verhalten vom Speicherbetreiber sichergestellt werden kann. So darf bei Frequenzen im Totband über 50Hz nur negative PRL und entsprechend unter 50 Hz nur positive PRL erbracht werden. Von zentraler Bedeutung ist dabei die Genauigkeit der Frequenzmessung: Es darf durch Messungenauigkeiten keinesfalls die Systemstabilität beeinträchtigt werden, so dass um 50 Hz herum ein ihnen entsprechender Puffer gebildet wird, in dem keine Regelleistung erbracht werden darf.[150] Nach einer von den ÜNBs 50Hertz, Tennet und TransnetBW eingeholten Auskunft, richtet sich die Erbringung der PRL im Totband, wie bei einer regulären PRL-Aktivierung, linear nach der Frequenzabweichung. Während Untererbringungen im Totband erlaubt sind, dürfen Übererbringungen nicht erfolgen, da sie eine PRL-Aktivierung in entgegengesetzter Richtung auslösen könnten.[151] Von dieser zusätzlichen Möglichkeit des Lademangagements soll auch in der Wirtschaftlichkeitsberechnung des fiktiven BSS Gebrauch gemacht werden.

Diese Ladeverfahren sollen nun in Abhängigkeit von der PRL-Aktivierung und unter Berücksichtigung der Vorgaben der ÜNB, bezüglich des Arbeitsbereiches des Ladezustandes, auf verschiedene Dimensionierungen des fiktiven BSS angewendet werden. Ein entscheidender Faktor ist dabei auch der Wirkungsgrad des fiktiven BSS. So verstärken die beim Einspeichern von Energie auftretenden Verluste den vorhandenen Überhang der positiven Regelleistung (vgl. Kapitel 5.3) und erhöhen außerdem, entsprechend ihres Ausmaßes, die über Fahrplangeschäfte zu beschaffende Energiemenge zum Ausgleich des Ladezustands auf seinen Sollwert. Koller et al. haben bei einem auf Lihtium-Ionen basierenden BSS einen durchschnittlichen Gesamtwirkungsgrad von 85% festgestellt, wobei 5% der Verluste auf das BMS zurückzuführen sind. Der exakte Gesamtwirkungsgrad hängt vor allem vom Leistungsniveau ab, auf dem das BSS betrieben wird. Seine Ermittlung ist komplex, da die Lithium-Ionen-Batterie besonders effizient bei niedrigen Leistungen arbeitet, während das BMS jedoch bei hohen Leistungen seine höchste Effizienz aufweist.[152] Auf der Bleisäure-Technologie basierende BSS haben laut Fuchs et al. einen durchschnittlichen Wirkungsgrad von 77,5%.[153] Die genannten Effizienzwerte sollen für die entsprechenden Varianten des fiktiven BSS übernommen werden und komplett mit den Einspeichervorgängen durch PRL-Aktivierung, Totbandladung oder Ausgleichsladungen, verrechnet werden. Für die Entladevorgänge wird im Gegenzug ein Wirkungsgrad von 100% angenommen.

Aus den Regelungen der ÜNB ergibt sich für ein Lithium-Ionen BSS mit 1,25 MWh Speicherkapazität und einer angebotenen Leistung von 1 MW ein Ladungszustand von 60%

[150] Deutsche ÜNB (2014b, S. 4–5)

[151] vgl. Online-Appendix A12 - 50Hertz Transmissions GmbH, Online-Appendix A13 - Tennet TSO GmbH, Online-Appendix A14 - TransnetBW GmbH

[152] Koller et al. (2015, S. 130–131)

[153] Fuchs et al. (2012, S. 44)

für die obere Grenze und ein SoC von 40% für die untere Grenze des Arbeitsbereichs. (vgl. Kapitel 4.5) Sobald einer der beiden Grenzwerte erreicht ist, wird der Ladungszustand durch Beschaffung von 125 kWh auf seinen Sollwert von 50% ausgeglichen. Wie in der Realität wurde bei der Anforderung der Ausgleichsenergie eine Zeitverzögerung von 15 Minuten implementiert. Der Ladestandsausgleich erfolgt mit den dafür vorgesehenen Leistungspuffer in Höhe von 25% der angebotenen Leistung (vgl. Kapitel 4.5). Somit vergehen von der Anforderung des Fahrplangeschäftes bis zur Vollendung der Ausgleichsladung exakt 45 Minuten. Für das Jahr 2013 müssten für die 631 durchgeführten Einspeichervorgänge insgesamt 92,79 MWh, beispielsweise über den Spot-Markt der Strombörse, beschafft werden. Der entsprechende Strompreis lag im Juni 2015 bei 3 ct/kWh.[154] Im Gegenzug werden zum Ausgleich des SoC auf 50% auch 30 MWh aus dem BSS entladen, die sich auf 240 Vorgänge verteilen. Unter der Annahme, dass die durch das Lademanagement ausgespeicherte Energie zu denselben Konditionen verkauft werden kann wie die für das Nachladen des BSS zu beschaffende Energie, muss für die Wirtschaftlichkeitsberechnung des BSS nur der Überhang der eingespeicherten im Vergleich zur ausgespeicherten Energie kostentechnisch erfasst werden. Mit einer Erhöhung des KLV steigt auch die Menge an Ausgleichsenergie, die bei Erreichen einer Grenze des Arbeitsbereiches angefordert wird. Da aber gleichzeitig die Zahl der Einspeicher- und Auspeichervorgänge deutlich zurückgeht, ergibt sich für das ganze Jahr ein verringerter Bedarf an Ausgleichsenergie. So sinkt bei einer Auslegungsgröße von 1,5 MWh die durch das Lademanagement eingespeicherte Energiemenge um 14,6% auf 79,4 MWh.[155] Bei Blei-Gel-BSS sind auf Grund des niedrigeren Wirkungsgrades im Vergleich zur Lithium-Ionen-Technologie bei gleichem KLV auf das Jahr gesehen mehr Einspeichervorgänge und weniger Entladungsvorgänge notwendig. So ergibt sich bei einer Auslegungsgröße von 1,25 MWh ein Saldo zu Gunsten der Einspeicherung von 83,99 MWh, im Vergleich zu 62,79 MWh bei einem Lithium-Ionen-BSS. Für die Ermittlung der Lebensdauer der verschiedenen BSS-Varianten im folgenden Kapitel ist von Bedeutung, wie viele äquivalente Vollzyklen das BSS durch das Lademanagement durchläuft. Die einzelnen Ausgleichsvorgänge haben, bei eben verwendeten KLV, eine Zyklentiefe von 10% der Speicherkapazität des BSS und sind in der Summe äquivalent zu 43,55 Vollzyklen bzw. 43,7 Vollzyklen bei einem Blei-Gel-BSS. Weitere 13,66 bzw. 12,82 Vollzyklen werden durch die Totbandladung verursacht.[156]

Durch die bei BSS übliche zeitabhängige Selbstentladung erhöht sich die zu beschaffende Menge Ausgleichsenergie. Laut den Angaben von Fuchs et al. beträgt die Selbstentladung bei der Lithium-Ionen-BSS 5% und bei der Bleisäure-BSS 12% der gespeicherten Energie pro Monat.[157] Der Wert für die Bleisäuretechnologie soll dabei für das Blei-Gel-BSS übernommen werden. Als Berechnungsbasis für die Verluste durch Selbstentladung soll die Energiemenge dienen, die dem Zielladezustand von 50% entspricht. Die entsprechenden Werte sind zusammen mit anderen bereits erläuterten Ergebnissen in Tabelle 6 zusammengefasst. Die aus der Selbstentladung resultierenden zusätzlichen Ladevorgänge sollen in Form von äquivalenten Vollzyklen bei der im folgenden Kapitel beschriebenen Ermittlung der Lebensdauern der verschiedenen Varianten des fiktiven BSS berücksichtigt werden.

5.5. Erwartete Lebensdauer

Für die Wirtschaftlichkeitsberechnung des fiktiven BSS soll eine Projektlaufzeit von 20 Jahren angenommen werden. In diesem Kapitel soll überprüft werden ob die zyklische und kalendarische Lebensdauer der verwendeten Batterietechnologien für die gesamte Projektlaufzeit ausreichend ist oder ob gegebenenfalls eine Zweitinvestition nötig wird. In den vorherigen Kapiteln wurden dazu bereits für die PRL-Aktivierung, das Lademanagement, die Totbandladung und die Selbstentladung die jeweilige Anzahl äquivalenter Vollzyklen bestimmt. Diese Einzelwerte werden nun zu einer Jahressumme aufaddiert und auf die Projektlaufzeit hochgerechnet. Für ein Lithium-Ionen-BSS mit einer Auslegungsgröße von 1,25 MWh ergeben sich so 2.426,56 äquivalente Vollzyklen die innerhalb von 20 Jahren durchlaufen werden müssen (vgl. Tabelle 7).

Laut einer Auskunft der Siemens AG können Lithium-Ionen-BSS mit NMC als Kathodenmaterial über ihre Lebensdauer ungefähr 5.000 Vollzyklen erbringen.[158] Da die für das fiktive BSS verwendete LFP-Technologie besonders langzeitstabil sein soll (vgl. Kapitel 3.2)[159], kann hier ein mindestens ebenso großer Wert erwartet werden. Die Zyklenlebensdauer wäre durch die 20-jährige Bereitstellung von PRL also bei weitem noch nicht ausgereizt. Die kalendarische Lebensdauer von Lithium-Ionen-BSS liegt laut Fuchs et al. bei bis zu 20 Jahren. Ein Erreichen dieses Wertes ist unter anderem vom durchschnittlichen SoC abhängig.[160] Da der Zielladezustand des fiktiven BSS von 50% sehr vorteilhaft für die Lebensdauer von Lithium-Ionen Batterien ist (vgl. Kapitel 3.2), soll für die weitere Wirtschaftlichkeitsberechnung eine kalendarische Lebensdauer von 20 Jahre für das Lithium-Ionen-BSS angenommen werden. Das auf der Blei-Gel-Technologie basierende BSS mit 1,25 MWh Auslegungsgröße kann mit einer zu erwarteten Zyklenlebensdauer von 2.900 äquivalenten Vollzyklen ebenfalls die Zyklisierungsanforderungen einer 20 jährigen Bereitstellung von PRL erfüllen. Die kalendarische Lebensdauer beträgt laut Hersteller jedoch nur 15 Jahre.[161] Da ein mittlerer Ladezustand von 50% zudem ungünstig für die Lebensdauer von Bleibatterien ist (vgl. Kapitel 2.1), kann mit einer Nutzung über die angegebene kalendarische Lebensdauer hinaus nicht gerechnet werden. Somit wird

[154] Internationales Wirtschaftsforum Regenerative Energien (2015)
[155] vgl. Online-Appendix A9 - Ladeverfahren
[156] vgl. Online-Appendix A9 - Ladeverfahren
[157] Fuchs et al. (2012, S. 42–44)
[158] vgl. Online-Appendix A5 - Siemens AG
[159] vgl. Online-Appendix A6 - Herstellerangabe
[160] Fuchs et al. (2012, S. 42)
[161] vgl. Online-Appendix A4 - Exide Technologies GmbH

Tabelle 6: Ausgleichsladung, Totbandladung und Selbstentladung

	Auslegungsgröße (MWh)							
	1,125	**1,25**	**1,375**	**1,5**	**1,125**	**1,25**	**1,375**	**1,5**
obere Grenze Arbeitsbereich	55,56%	60,00%	63,64%	66,67%	55,56%	60,00%	63,64%	66,67%
untere Grenze Arbeitsbereich	44,44%	40,00%	36,36%	33,33%	44,44%	40,00%	36,36%	33,33%
	Lithium-Ionen (LFP)				**Blei-Gel**			
Einspeichermenge Ausgleich (MWh)	130,59	92,79	84,93	79,41	148,31	108,87	100,65	96,45
Ausspeichermenge Ausgleich (MWh)	-58,50	-30,00	-25,31	-21,50	-49,88	-24,88	-18,75	-16,50
Einspeichermenge Totban (MWh)	19,99	21,57	22,92	23,53	18,95	21,31	21,72	22,27
Ausspeichermenge Totband (MWh)	-14,63	-12,58	-11,81	-11,54	-14,05	-10,74	-10,95	-10,45
Selbstentladung pro Jahr (MWh)	337,50	375,00	412,50	450,00	810,00	900,00	990,00	1080,00

Tabelle 7: Zyklenanforderungen an und zyklische Lebensdauer von Lithium-Ionen (LFP) und Blei-Gel BSS

		Auslegungsgröße (MWh)			
		1125	**1250**	**1375**	**1500**
Lithium-Ionen (LFP)	jährliche Zyklenanforderung	170,10	128,67	113,06	100,94
	Zyklenanforderung 20 Jahre	3402,03	2573,44	2261,29	2018,80
	Zyklenlebensdauer	5000,00	5000,00	5000,00	5000,00
	zyklisches Lebensende (Jahren)	29,39	38,86	44,22	49,53
Blei-Gel	jährliche Zyklenanforderung	167,59	128,27	112,33	101,19
	Zyklenanforderung 20 Jahre	3351,79	2565,44	2246,61	2023,88
	Zyklenlebensdauer	2900,00	2900,00	2900,00	2900,00
	zyklisches Lebensende (Jahren)	17,30	22,61	25,82	28,66

bei dem Blei-Gel-BSS nach 15 Jahren eine Ersatzinvestition in die Batterietechnik notwendig. Am Ende der Projektlaufzeit von 20 Jahren soll der Restwert der Blei-Gel-Batterie als Liquidationserlös verbucht werden.

5.6. Dimensionierung des BSS

Das Kapitel Dimensionierung dreht sich nicht nur um die absoluten Werte von Nennleistung und Speicherkapazität, sondern vor allem auch um das KLV, einem Parameter, dem bei der Bereitstellung von PRL eine große Bedeutung zukommt. Durch die allgemein für PRL geltende Mindestangebotsgröße von 1 MW (vgl. Kapitel 2.8.3) und dem speziell für BSS geltenden Leistungspuffer zur Nutzung für Ausgleichsenergie in Höhe von 25% der am PRL-Markt präqualifizierten Leistung (vgl. Kapitel 4.5), ergibt sich eine Mindestnennleistung von 1,25 MW. Diese wurde in Kapitel 5.1 bereits als Nennleistung des fiktiven BSS festgesetzt. Die Speicherkapazität soll im Einklang mit geltenden Regelungen variiert werden, um so die Wirtschaftlichkeit verschiedener KLV prüfen zu können. Nicht erlaubt für BSS ist seit August 2015 beispielsweise ein Verhältnis von Speicherkapazität zu Angebotsleistung (KALV) von weniger als 1, da in diesem Fall keine volle Leistungsabrufung für 30 Minuten in beide Regelrichtungen gewährleistet werden kann (vgl. Kapitel 4.5). Für nicht speicherbegrenzte Anlagen gilt dagegen nach wie vor das vom Transmission Code 2007 vorgesehene Qualifizierungskriterium, einer 15 minütigen positiven bzw. negativen Vollaktivierung (vgl. Kapitel 2.8.1). Oudalov et al. haben 2007 unter Verwendung von Fahrplangeschäften, Totbandladung und der Anpassung von oberer und unterer Ladezustandsgrenze ein KALV von 0,62 als ideal für ein BSS bestimmt.[162] Eine solche Auslegung des BSS ist heutzutage nicht mehr regelkonform. Auch ein ausgeglichenes Verhältnis von angebotener Leistung und Speicherkapazität soll für das fiktive BSS ebenfalls nicht weiter verfolgt werden. Während diese Variante theoretisch erlaubt ist, kann sie in der Praxis leider nicht regelkonform umgesetzt werden. Sie hätte eine obere und untere Grenze des Arbeitsbereiches von 50% zur Folge, so dass selbst kleinste Abrufungen oder Ausgleichsladungen, die minimal über den Zielladezustand von 50% hinaus gehen, zu einem Verlassen des Arbeitsbereiches führen würden, was nur bei Frequenzabweichungen größer 50 mHz toleriert wird (vgl. Kapitel 4.5). Letztlich sollen für die zwei gewählten Batterietechnologien jeweils vier Auslegungsvarianten mit einem KALV zwischen 1,125 und 1,5 in der Profitabilitätsrechnung im folgenden Kapitel berücksichtigt werden. Eine weitere wichtige Implikation für die Dimensionierung des fiktiven BSS stammt aus dem Alterungsverhalten der Batterieeinheiten. Mit der Zeit verlieren diese kontinuierlich an Speichervermögen. Das Ende ihrer Lebensdauer gilt gemeinhin als erreicht wenn, sie nur noch 80% der ursprünglichen Speicherkapazität aufweisen (vgl Kapitel 3). Die für das fikti-

[162] Oudalov et al. (2007, S. 1266)

ve BSS gewählten Batterietechnologien erreichen diesen Zustand nach 20 bzw. 15 Betriebsjahren (vgl. Kapitel 5.5). Da zu diesem Zeitpunkt die Bereitstellung von PRL genauso sichergestellt werden soll wie am ersten Betriebstag, wird ein Puffer in Höhe von 20% der Speicherkapazität in die Investitionskosten mit einkalkuliert. Bei den anderen Inputfaktoren der Wirtschaftlichkeitsberechnung wird der Puffer nicht berücksichtigt, da er nicht an der Bereitstellung von PRL teilnimmt, sondern nur als eine Art "kalte Reserve" verlorengegangene Speicherkapazität ersetzt. Die vier untersuchten Auslegungsvarianten sind samt ihrer Puffergröße in Tabelle 8 dargestellt.

5.7. Kapitalwertberechnung

Zur Bestimmung der Wirtschaftlichkeit eines BSS, das über einen Zeitraum von 20 Jahren PRL bereitstellt, soll die Kapitalwertmethode (siehe Formel 5) verwendet werden.[163]

$$K_0 = \sum_{t=0}^{n} \frac{e_t - a_t}{(1+i)^t} + \frac{L_n}{(1+i)^n} - I_0 \qquad (5)$$

Formel zur Berechnung des Kapitalwerts

Anhand des fiktiven Blei-Gel-BSS mit einer Auslegungsgröße von 1,25 MWh (KLV = 1) soll dargestellt werden, in welche Variablen der Kapitalwertgleichung die in den vorangegangen Kapiteln ermittelten Werte eingesetzt werden. Alle eingesetzten Werte sind übersichtlich in Online-Appendix A11 - Kapitalwert dargestellt. In den Einzahlungen e_t wird jeweils der Jahresertrag aus den bei der PRL-Ausschreibung akzeptierten Geboten eingefügt. Mit der verfolgten Strategie, die für beide Batterietechnologien und alle Auslegungsvarianten identisch ist, wird jährlich ein Betrag von 150.327,84 Euro erwirtschaftet (vgl. Kapitel 5.2). Die Auszahlungen a_t setzen sich aus den Betriebskosten von jährlich 20.000 Euro (vgl. Kapitel 5.1) und den Kosten für die Ausgleichsenergie und die Selbstentladung zusammen (vgl. Kapitel 5.4). Da für den gesamten Projektzeitraum die Werte des Jahres 2013 verwendet werden, haben die Kosten für die Energiebeschaffung in jedem Jahr eine Höhe von 2.554,72 Euro. Die Auszahlungen für Ausgleichsenergie und Selbstentladung sind somit kleiner als 2% des Umsatzes durch die PRL-Erbringungen. In der Periode 15 werden den Auszahlungen a_t einmalig die Kosten für die nötige Ersatzinvestition für die Blei-Gel-Batterietechnik hinzugefügt (vgl. Kapitel 5.5). Da laut den Angaben des Herstellers Exide Technologies GmbH, die Batterie einen Anteil von 50% an den Investitionskosten des BSS hat, wird für die Ersatzinvestition der entsprechende Wert von 712.500 Euro angesetzt.[164] Da die erneuerte Batterietechnik am Ende der Projektlaufzeit erst ein Drittel ihrer kalendarischen Lebensdauer aufgebraucht hat, werden für den Liquidationserlös L_n zwei Drittel des Betrages der Ersatzinvestition eingesetzt. Für die Auszahlung I_0 werden die initialen Investitionskosten des BSS eingefügt. Zur Ermittlung von selbigen werden zunächst die spezifischen Investitionskosten in € /kWh für ein KLV von 1,2 benötigt. Dieser, im Vergleich zur gemachten Annahme, um 20% erhöhte Wert resultiert aus dem bei den Investitionskosten zu berücksichtigenden Puffer für die Speicherkapazität (vgl. Kapitel 5.6). Die spezifischen Investitionskosten können mit Hilfe der in Kapitel 5.1 ermittelten Formel 5.1 errechnet werden und betragen 950 € /kWh. Durch Multiplikation mit der Speicherkapazität inklusive Puffer von 1,5 MWh ergeben sich die initialen Investitionskosten in Höhe von 1.425.000 Euro. Nachdem für die Nutzungsdauer in Jahren n der Wert 20 eingesetzt wird, fehlt für eine Berechnung des Kapitalwerts nur noch der Diskontierungszinsatz i. Hierfür werden der von KPMG in der Kapitalkostenstudie 2014 veröffentlichte gewichtete durchschnittliche Kapitalkostensatz (WACC) der Energiebranche in Höhe von 7,2% verwendet.[165] Die Ergebnisse der Kapitalwertberechnung für die jeweils vier untersuchten Auslegungsvarianten der beiden Batterietechnologien sind in Tabelle 9 dargestellt.[166]

Aus den negativen Kapitalwerten für alle untersuchten Varianten kann interpretiert werden, dass eine Investition in ein BSS zur Bereitstellung von PRL unter den gegebenen Bedingungen nicht empfehlenswert ist. Bei dem Vergleich der beiden Batterietechnologien schneiden die Varianten mit Blei-Gel-Batterien, trotz der nötigen Ersatzinvestition, besser ab als die konkurrierenden Lithium-Ionen Batterien. Dies liegt vor allem an den günstigeren fixen Betriebskosten, die je nach Auslegungsvariante zwischen 41 und 46 Prozent niedriger liegen. Stellt man die verschiedenen Auslegungsgrößen in Bezug zueinander, wird deutlich, dass, wie erwartet, der Kapitalwert steigt, wenn die Speicherkapazität reduziert wird (vgl. Kapitel 5). Für einen Break Even des Blei-Gel BSS mit 1,25 MWh Speicherkapazität, müssten dessen spezifische Investitionskosten von derzeit 950 € /kWh auf 799,99 € /kWh fallen.[167] Eine Möglichkeit zur Erhöhung des Kapitalwertes aller untersuchten Varianten ist der Verzicht auf die initiale Anschaffung des Puffers für die Speicherkapazität. Unter dieser Bedingung ergibt sich für ein Blei-Gel-BSS mit einer Auslegungsgröße von 1,125 MWh ein positiver Kapitalwert in Höhe von 62.316,23 Euro. Alle anderen Varianten weisen jedoch weiterhin einen negativen Kapitalwert auf.[168] Um die Bereitstellung von PRL trotz fehlenden Puffers sicherstellen zu können, muss der Vertreiber der Batterietechnik dazu in Lage sein, während der Projektlaufzeit, je nach den aktuell aufgetretenen Verlusten, Batteriezellen mit entsprechender Speicherkapazität nachzuliefern. Dabei entstehen wiederum Kosten, die den Kapitalwert verringern, allerdings fallen diese wegen der erfolgenden Diskontierung weniger stark ins Gewicht als die initiale Beschaffung eines Puffers für die Speicherkapazität. Außerdem wirken sich bei der verteilten Beschaffung die für die Zukunft erwarteten niedrigeren Batteriepreise aus. So sollen laut Fuchs et al. für eine Auslegung

[163] Thommen und Achleitner (2009, S. 707)
[164] vgl. Online-Appendix A4 - Exide Technologies GmbH
[165] KPMG (2014)
[166] vgl. Online-Appendix A11 - Kapitalwertberechnung
[167] vgl. Online-Appendix A11 - Kapitalwertberechnung
[168] vgl. Online-Appendix A11 - Kapitalwertberechnung

Tabelle 8: Dimensionierung der Auslegungsvarianten für die Wirtschaftlichkeitsberechnung

Nennleistung (MW)	1,25	1,25	1,25	1,25
angebotene Leistung (MW)	1	1	1	1
Speicherkapazität (MWh)	1,125	1,25	1,375	1,5
Speicherkapazität inkl. Puffer (MWh)	1,35	1,5	1,65	1,8
KLV	0,9	1	1,1	1,2
KALV	1,125	1,25	1,375	1,5
KLV (inkl. Puffer)	1,08	1,2	1,32	1,44

Tabelle 9: Kapitalwert und spezifische Break Even Kosten der Auslegungsvarianten

	Auslegungsgröße (MWh) (exkl. Puffer)			
	1,125	**1,25**	**1,375**	**1,5**
		Kapitalwert (€)		
Lithium-Ionen (LFP)	-291.570,79€	-460.897,98€	-625.760,53€	-784.707,65€
Blei-Gel	-107.558,77€	-225.019,00€	-336.500,11€	-438.191,32€
		Break Even Investitionskosten (€/kWh)		
Lithium-Ionen (LFP)	897,02€	787,73€	697,75€	623,05€
Blei-Gel	900,33€	799,99€	716,06€	646,56€

von 1,25 MW Nennleistung und 1,25 MW Speicherkapazität bis zum Jahr 2030, ausgehend von den Werten des Jahres 2012, die Kosten für Bleibatterien um rund 67% sinken und die Preise von Lithium-Ionen-Batterien um ca. 62% fallen. [169] Diese Preisentwicklung hätte auch einen zusätzlichen positiven Effekt speziell auf die Kapitalwerte der Auslegungsvarianten mit Blei-Gel-Batterien, da die Kosten für die in Periode 15 nötige Ersatzinvestition sinken würden.

In anderen Ländern können Batteriespeicher bereits profitabel PRL erbringen. Eine Untersuchung für den dänischen Markt kommt zum Ergebnis, dass die Bereitstellung von PRL mit einem 1 MWh/1 MW großen Lithium-Ionen-BSS, mit LFP und Graphit als Aktivmaterialien, profitabel ist. Bei einer Laufzeit von 20 Jahren kann ein Kapitalwert von 10.803.651 Euro erzielt werden. [170] Allerdings unterscheiden sich dort auch die Marktvoraussetzungen für PRL deutlich von denen in Deutschland (vgl. Kapitel 2.9). Eine Möglichkeit, die Profitabilität von BSS bei der Bereitstellung von PRL zu erhöhen, wäre folglich die Änderung der hiesigen Bedingungen am Primärregelmarkt.

So schlägt Prof. Weyer in seinem Beitrag zum 3. Energiepolitischen Workshop vor, das Vergütungssystem für die Erbringung von Regelleistung dahingehend zu ändern, dass auch Schnelligkeit und Genauigkeit der Erbringung in der Bepreisung berücksichtigt werden. [171] Da Batterien ihre volle Leistung innerhalb von Millisekunden abrufen und diese beliebig innerhalb ihrer Nennleistung regeln können (vgl. Kapitel 4.2), würden sich durch die von Prof. Weyer vorgeschlagenen Änderungen die Gewinnaussichten von Betreibern von Batteriespeichern für die Bereitstellung von PRL verbessern. Darüber hinaus kritisiert Prof. Weyer auch, dass Batteriespeicher rein rechtlich bisher als Letzverbraucher behandelt werden. Dies entspreche nicht der rechtlichen Definition von Letztverbrauchern nach § 3 Nr. 25 EnWG[172], die besagt, dass Letzverbraucher "natürliche oder juristische Personen" sind, "die Energie für den eigenen Verbrauch kaufen".[173] Eine Befreiung der Batteriespeicher von Status des Letzverbrauchers hätte eine finanzielle Entlastung für diese zur Folge, da dann Abgaben wie Netzentgelte, Stromsteuer und Konzessionsabgaben nicht mehr zu entrichten wären. Eine dementsprechende Änderung der Rechtslage ist nicht unwahrscheinlich, da laut Koalitionsvertrag der aktuellen Bundesregierung der Status der Stromspeicher als Letztverbraucher auf den Prüfstand gestellt werden soll.[174] So soll laut einem Plan des Bundesministeriums für Wirtschaft und Energie Ende 2015/Anfang 2016 ein neues Gesetz zum Strommarktdesign, eine Novelle des EnWG, auf den Weg gebracht werden. [175] Vielleicht wird dieses schon neue Implikationen für den Einsatz von Batteriespeichern zur Regelenergiebereitstellung beinhalten.

6. Konklusion und Ausblick

Abschließend bleibt festzuhalten, dass Lithium-Ionen BSS, mit ihrer hohen zyklischen und kalendarischen Lebensdauer, ihrem ausgezeichneten Wirkungsgrad und den für sie

[169] Fuchs et al. (2012, S. 42, 44)
[170] Swierczynski et al. (2013, S. 491–492)
[171] Weyer (2014, S. 21)
[172] Weyer (2014, S. 8–12)
[173] Deutscher Bundestag (2005)
[174] Weyer (2014, S. 8–12)
[175] Bundesministerium für Wirtschaft und Energie (2015)

idealen Ladungsständen zwischen 40% und 70% aus technischer Sicht die am Besten geeignete Batterietechnologie für die Bereitstellung von PRL sind (vgl. Kapitel 3.2, Kapitel 4.4). Die auf der länger etablierten Blei-Säure-Technologie basierenden BSS haben dafür mit ihren vergleichsweise geringen Investitions- und Betriebskosten in der Wirtschaftlichkeitsberechnung besser abschneiden können.(vgl. Kapitel 5.7) Um jedoch profitabel betrieben zu werden müssen jedoch entweder die Eigenschaften der beiden Batterietechnologien weiter verbessert werden oder aber günstigere Rahmenbedingungen für den Einsatz von BSS zur Bereitstellung von PRL geschaffen werden.

Ein ganz anderes Konzept sieht die Zukunft der PRL-Erbringung auf Seiten der Verbraucher. So ist es vorstellbar durch die Nutzung des sog. Smart Grid, Leistungsungleichgewichte, die zur Aktivierung von PRL führen, erst gar nicht aufkommen zu lassen. Zu Spitzenlastzeiten könnten weniger dringliche Verbraucher, wie Kühlschränke und Klimaanlagen, kurzzeitig vom Netz getrennt werden, um so Lastspitzen zu vermeiden und das Netz insgesamt zu stabilisieren. In Zusammenarbeit mit Energiespeichersystemen kann dieses System eine noch größere Wirkung entfalten und so ein sehr effektives Lastmanagement verwirklichen.[176] Auch in diesem Konzept können Energiespeicher wie die Batteriespeichersysteme also einen positiven Beitrag zur Netzstabilisierung leisten.

Daher ist es naheliegend, dass in die Erforschung neuer revolutionärer Batterietechnologien investiert wird. So fördert das Bundesministerium für Wirtschaft und Energie (BMWi) zur Zeit die Entwicklung von Zink-Luft-Batteriesystemen für den Einsatz in stationären Energiespeichern im Megawatt-Bereich. Diese Technologie verspricht beeindruckende spezifische Energiedichten von bis zu 1100 Wh/kg. Als Hauptproblem stellt sich bisher die geringe Zyklenlebensdauer dar. Nach Abschluss einer ersten Projektstufe im August 2015 soll nun eine Laboranlage errichtet werden, die einen Wirkungsgrad von 60% und eine Lebensdauer von mindestens 500 Zyklen erreicht. [177] Eine Marktreife hat diese Technologie zwar, in Anbetracht von Lithium-Ionen-Batterien mit Wirkungsgraden von 85% und einer nutzungsbezogenen Lebensdauer von bis zu 5000 Zyklen (vgl. Kapitel 5.4, 5.5), noch nicht erreicht. In der hohen Energiedichte steckt jedoch viel Potential, sodass das BMWi mit der Zink-Luft-Technologie vielleicht eine Technologie fördert, die mittelfristig zu einem wichtigen Baustein für das Gelingen der Energiewende werden kann.

[176] Lindley (2010, S. 20)
[177] Energiespeicher - Forschungsinitiative der Bundesregierung (2015)

Literatur

Anderson, M. und Carr, D. Battery energy storage technologies. *Proceedings of the IEEE*, 81(3):475–479, 1993. ISSN 00189219. doi: 10.1109/5.241482.

Birke, P. und Schiemann, M. *Akkumulatoren: Vergangenheit, Gegenwart und Zukunft elektrochemischer Energiespeicher*. Utz, München, 2013. ISBN 978-3-8316-0958-1.

Boje, E., Camisani-Calzolari, F., und Borsche, T. Impact of frequency control reserve provision by storage systems on power system operation. In Boje, E., editor, *World Congress*, IFAC proceedings volumes, pages 4038–4043. IFAC, Elsevier, 2014a. doi: 10.3182/20140824-6-ZA-1003.02257.

Boje, E., Camisani-Calzolari, F., und Ulbig, A. Impact of low rotational inertia on power system stability and operation. In Boje, E., editor, *World Congress*, IFAC proceedings volumes, pages 7290–7301. IFAC, Elsevier, 2014b. doi: 10.3182/20140824-6-ZA-1003.02615.

Borsche, T., Ulbig, A., Koller, M., und Andersson, G. Power and energy capacity requirements of storages providing frequency control reserves. In *2013 IEEE Power & Energy Society General Meeting*, pages 1–5, 2013. doi: 10.1109/PESMG.2013.6672843.

Bundesministerium für Wirtschaft und Energie. Energie der Zukunft: Die Energiewende geht weiter, 2015. URL http://www.bmwi.de/DE/Themen/Energie/Energiewende/gesamtstrategie.html. (besucht am 09. 10. 2015).

Bundesnetzagentur. Festlegung zu Verfahren zur Ausschreibung von Regelenergie in Gestalt der Primärregelung: BK6-06-065, 2007. URL http://www.bundesnetzagentur.de/DE/Service-Funktionen/Beschlusskammern/1BK-Geschaeftszeichen-Datenbank/BK6-GZ/2006/2006_001bis099/BK6-06-065/BK6-06-065_Beschluss_pdf.pdf?__blob=publicationFile&v=4. (besucht am 08. 10. 2015).

Bundesnetzagentur. Bundesnetzagentur ordnet Netzregelverbund für die deutschen Stromnetze an: Kurth: „Kosteneinsparungen in dreistelliger Millionenhöhe zu erwarten“, 2010. URL http://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Allgemeines/Presse/Pressemitteilungen/2010/100316NetzregelverbundStromId18377pdf.pdf;jsessionid=C75499D234D111BAE69BB870B47A69CB?__blob=publicationFile&v=3. (besucht am 11. 10. 2015).

Bundesnetzagentur. Festlegung zu Verfahren zur Ausschreibung von Regelenergie in Gestalt der Primärregelung: Beschluss BK6-10-097, 2011. URL http://www.bundesnetzagentur.de/DE/Service-Funktionen/Beschlusskammern/1BK-Geschaeftszeichen-Datenbank/BK6-GZ/2010/2010_001bis099/BK6-10-097bis-099/BK6-10-097_Beschluss_2011_04_12.pdf?__blob=publicationFile&v=3. (besucht am 09. 10. 2015).

Bundesnetzagentur. Szenariorahmen 2025: Genehmigung: Az.: 6.00.03.05/14-12-19/szenariorahmen 2025, 2014. URL http://www.netzausbau.de/SharedDocs/Downloads/DE/Delta/Szenariorahmen/Szenariorahmen_2025_Genehmigung.pdf?__blob=publicationFile. (besucht am 07. 10. 2015).

Consentec GmbH. Beschreibung von Regelleistungskonzepten und Regelleistungsmarkt: Studie im Auftrag der deutschen Übertragungsnetzbetreiber: 50Hertz Transmission GmbH, 2014.

Deutsche ÜNB. Information zum netzregelverbund und der internationalen weiterentwicklung: Beitritt österreich (apg) zum internationalen netzregelverbund (igcc – international grid control cooperation), 2014a. URL https://www.regelleistung.net/ext/download/marktinformationenApg.

Deutsche ÜNB. Eckpunkte und Freiheitsgrade bei Erbringung von Primärregelleistung, 2014b. URL https://www.regelleistung.net/ext/download/eckpunktePRL. (besucht am 12. 10. 2015).

Deutsche ÜNB. Anforderungen an die Speicherkapazität bei Batterien für die Primärregelleistung, 2015. URL https://www.regelleistung.net/ext/download/anforderungBatterien. (besucht am 08. 10. 2015).

Deutscher Bundestag. Gesetz über die elektrizitäts- und gasversorgung: Energiewirtschaftsgesetz - enwg, 2005.

Deutscher Bundestag. Energiekonzept für eine umweltschonende, zuverlässige und bezahlbare Energieversorgung: Drucksache 17/3049, 2010. URL http://dipbt.bundestag.de/doc/btd/17/030/1703049.pdf. (besucht am 11. 10. 2015).

Divya, K. und Østergaard, J. Battery energy storage technology for power systems—an overview. *Electric Power Systems Research*, 79(4):511–520, 2009. ISSN 03787796. doi: 10.1016/j.epsr.2008.09.017.

Doughty, D., Butler, P., Akhil, A., Clark, N., und Boyes, J. Batteries for large-scale stationary electrical energy storage: Electrochemical society interface. 19(3):49–53, 2010.

Elkraft System and Eltra Regulation. Wind turbines connected to grids with voltages above 100 kv: -technical regulation for the properties and the regulation of wind turbines„ 2004. URL http://www.energinet.dk/SiteCollectionDocuments/Engelske%20dokumenter/El/Grid%20Code%203.2.5%20Wind%20Turbines%20connected%20above%20100%20kV.pdf. (besucht am 30. 09. 2015).

Energiespeicher - Forschungsinitiative der Bundesregierung. Zink-Luft-Akkus fürs Netz, 2015. URL http://forschung-energiespeicher.info/projektschau/gesamtliste/projekt-einzelansicht/95/Zink_Luft_Akkus_fuers_Netz/.

ENTSO-E - European Network of Transmission System Operators for Electricity. Continental europe operation handbook - p1: Policy 1: Load-frequency control - final version, 2009.

Flinkerbusch, K. und Heuterkes, M. Cost reduction potentials in the german market for balancing power. *Energy Policy*, 38(8):4712–4718, 2010. ISSN 03014215. doi: 10.1016/j.enpol.2010.04.038.

Fraunhofer-Institut für Windenergie und Energiesystemtechnik, Institute für Elektrische Anlagen und Energiewirtschaft, und Stiftung für Umweltenergierecht. Roadmap Speicher: Bestimmung des Speicherbedarfs in Deutschland im europäischen Kontext und Ableitung von technisch-ökonomischen sowie rechtlichen Handlungsempfehlungen für die Speicherförderung, 2014. URL https://www.energie.fraunhofer.de/de/bildmaterial/news-pdf/roadmap-speicher.pdf. (besucht am 05. 10. 2015).

Fuchs, G., Lunz, B., Leuthold, M., und Sauer, D. Technology overview on electricity storage: Overview on the potential and on the deployment perspectives of electricity storage technologies, 2012. URL http://www.sefep.eu/activities/projects-studies/120628_Technology_Overview_Electricity_Storage_SEFEP_ISEA.pdf. (besucht am 14. 10. 2015).

Hudson, R., Kirby, R., und Wan, Y. Regulation requirements for wind generation facilities. In *Proceedings of the Windpower 2001 Conference*, pages 1–10, 2001. URL http://www.consultkirby.com/files/AWEA_Wind_Regulation.pdf.

Internationales Wirtschaftsforum Regenerative Energien. Strompreise für Verbraucher steigen - Börsen-Strompreise sinken auf Rekordtiefs, 2015. URL http://www.iwr-institut.de/de/presse/presseinfos-energiewende/strompreise-fuer-verbraucher-steigen-boersen-strompreise-sinken-auf-rekordtiefs. (besucht am 05. 10. 2015).

Jansen, M. Optimierung der Marktbedingungen für die Regelleistungserbringung durch Erneuerbare Energien, 2014. URL http://www.energiecluster.de/files/20140407_bee_optimierungdermarktbedingungenfrregelleistungserbringungfuererneuerbareenergien.pdf. (besucht am 09. 10. 2015).

Jin, C., Lu, N., Lu, S., Makarov, Y., und Dougal, R. Coordinated control algorithm for hybrid energy storage systems. In *Energy Society General Meeting*, pages 1–7, 2011. doi: 10.1109/PES.2011.6039893.

Joseph, A. und Shahidehpour, M. Battery storage systems in electric power systems. In *2006 IEEE Power Engineering Society General Meeting*, page 8 pp, 2006. doi: 10.1109/PES.2006.1709235.

Jossen, A. und Weydanz, W. *Moderne Akkumulatoren richtig einsetzen: 36 Tabellen*. Ubooks-Verl. and Reichardt, Neusäß and Untermeitingen, 2006. ISBN 978-3-939359-11-1.

Kempkens, W. Größter Batteriespeicher Europas geht in Schwerin ans Netz. Ingenieur.de, 2014a. URL http://www.ingenieur.de/Themen/Energiespeicher/Groesster-Batteriespeicher-Europas-geht-in-Schwerin-ans-Netz. (besucht am 12. 10. 2015).

Kempkens, W. Speicher für Grünstrom: Projekt für Großbatterie startet in Aachen, 2014b. URL http://green.wiwo.de/speicher-fuer-gruenstrom-projekt-fuer-grossbatterie-startet-in-aachen/. (besucht am 30. 09. 2015).

Koller, M., Borsche, T., Ulbig, A., und Andersson, G. Defining a degradation cost function for optimal control of a battery energy storage system. In *2013 IEEE Grenoble PowerTech*, pages 1–6, 2013. doi: 10.1109/PTC.2013.6652329.

Koller, M., Borsche, T., Ulbig, A., und Andersson, G. Review of grid applications with the zurich 1mw battery energy storage system. *Elec-*

tric Power Systems Research, 120:128–135, 2015. ISSN 03787796. doi: 10.1016/j.epsr.2014.06.023.

KPMG. Kapitalkostenstudie 2014, 2014. URL http://www.kpmg.com/DE/de/Documents/kapitalkostenstudie2014-de.pdf.

Kunisch, H., Kramer, K., und Dominik, H. Battery energy storage another option for load-frequency-control and instantaneous reserve. *IEEE Transactions on Energy Conversion*, EC-1(3):41–46, 1986. ISSN 0885-8969. doi: 10.1109/TEC.1986.4765732.

Lachs, W. und Sutanto, D. Application of battery energy storage in power systems. In *1995 International Conference on Power Electronics and Drive Systems. PEDS 95*, pages 700–705, 1995. doi: 10.1109/PEDS.1995.404984.

Lindley, D. Smart grids: The energy storage problem. *Nature*, 463(7277): 18–20, 2010. ISSN 0028-0836. doi: 10.1038/463018a.

Lippe, P. *Deskriptive Statistik*, volume 1632 of *Uni-Taschenbücher*. Fischer, Stuttgart [u.a.], 1993. ISBN 3-8252-1632-2.

Lo, C. und Anderson, M. Economic dispatch and optimal sizing of battery energy storage systems in utility load-leveling operations. *IEEE Transactions on Energy Conversion*, 14(3):824–829, 1999. ISSN 08858969. doi: 10.1109/60.790960.

Mercier, P., Cherkaoui, R., und Oudalov, A. Optimizing a battery energy storage system for frequency control application in an isolated power system. *IEEE Transactions on Power Systems*, 24(3):1469–1477, 2009. ISSN 0885-8950. doi: 10.1109/TPWRS.2009.2022997.

Oudalov, A., Chartouni, D., Ohler, C., und Linhofer, G. Value analysis of battery energy storage applications in power systems. In *2006 IEEE PES Power Systems Conference and Exposition*, pages 2206–2211, 2006. doi: 10.1109/PSCE.2006.296284.

Oudalov, A., Chartouni, D., und Ohler, C. Optimizing a battery energy storage system for primary frequency control. *IEEE Transactions on Power Systems*, 22(3):1259–1266, 2007. ISSN 0885-8950. doi: 10.1109/TPWRS.2007.901459.

Rebours, Y., Kirschen, D., Trotignon, M., und Rossignol, S. A Survey of Frequency and Voltage Control Ancillary Services—Part I: Technical Features. *IEEE Transactions on Power Systems*, 22(1):350–357, 2007a. ISSN 0885-8950. doi: 10.1109/TPWRS.2006.888963.

Rebours, Y., Kirschen, D., Trotignon, M., und Rossignol, S. A Survey of Frequency and Voltage Control Ancillary Services—Part II: Economic Features. *IEEE Transactions on Power Systems*, 22(1):358–366, 2007b. ISSN 0885-8950. doi: 10.1109/TPWRS.2006.888965.

Retzbach, L. *Akkus und Ladetechniken*. Franzis-Elektronik. Franzis, Poing, 2008. ISBN 978-3-7723-4170-0.

Rivero, E., Barquin, J., und Rouco, L. European balancing markets. In *2011 European Energy Market (EEM)*, pages 333–338, 2001. doi: 10.1109/EEM.2011.5953033.

Roberts, B. und McDowall, J. Commercial successes in power storage. *IEEE Power and Energy Magazine*, 3(2):24–30, 2005. ISSN 1540-7977. doi: 10.1109/MPAE.2005.1405867.

Scrosati, B. und Garche, J. Lithium batteries: Status, prospects and future. *Journal of Power Sources*, 195(9):2419–2430, 2010. ISSN 03787753. doi: 10.1016/j.jpowsour.2009.11.048.

Stöckl, G. *Integration der Elektromobilität in das Energieversorgungsnetz*. Elektrotechnik. Verl. Dr. Hut, München, 2014. ISBN 978-3-8439-1796-4.

Swierczynski, M., Stroe, D., Stan, A., und Teodorescu, R. Primary frequency regulation with li-ion battery energy storage system: A case study for Denmark. In *2013 IEEE ECCE Asia Downunder (ECCE Asia 2013)*, pages 487–492, 2013. doi: 10.1109/ECCE-Asia.2013.6579141.

Thommen, J.-P. und Achleitner, A.-K. *Allgemeine Betriebswirtschaftslehre: Umfassende Einführung aus managementorientierter Sicht*. Gabler, Wiesbaden, 2009. ISBN 978-3-8349-1325-8.

Verband der Netzbetreiber – VDN – e.V. beim VDEW. Transmissioncode 2003 - anhang d1: Unterlagen zur präqualifikation für die erbringung von primärregelleistung für die ünb, 2003. URL https://www.regelleistung.net/ext/download/praequalifikationD1. (besucht am 12. 10. 2015).

Verband der Netzbetreiber – VDN – e.V. beim VDEW. Transmissioncode 2007 - netz und systemregeln der deutschen übertragungsnetzbetreiber, 2007. URL https://www.bdew.de/internet.nsf/id/A2A0475F2FAE8F44C12578300047C92F/$file/TransmissionCode2007.pdf. (besucht am 07. 10. 2015).

Weissbach, T. und Welfonder, E. High frequency deviations within the european power system - origins and proposals for improvement. *VGB PowerTech*, (6/2009):26–34, 2009.

Weyer, H. Energiespeicherausbau –Regulative Herausforderungen: 3. Energiepolitischer Workshop „Energiespeicher – das Stiefkind der deutschen Energiewende“: Ifo-Institut, München, 16. September 2014, 2014. URL http://www.iber.tu-clausthal.de/fileadmin/IBER/Vortraege/140916_Vortrag_Stromspeicher_Ifo-Institut.pdf. (besucht am 13. 10. 2015).