JUMS

Volume 4, Issue 3, September 2019

JUNIOR MANAGEMENT SCIENCE

Belinda Kellerer, Portfolio Optimization and Ambiguity Aversion 305

Tobias Wulfert, Mobile App Service Quality Dimensions and Requirements for Mobile Shopping Companion Apps 339

Virginia Springer, Bewertung der Übertragbarkeit von neuronalen Studienergebnissen auf einen Accounting-Kontext 392

Lukas Ferner, Measuring the Impact of Carbon Emissions on Firm Value Using Quantile Regression 422

Leona Schink, Die Legitimation einer Innovation durch Cultural Entrepreneurship – Explorative Fallstudie eines symbiotischen Zusammenspiels zwischen einem Start-up und dessen Schlüsselkunden 433

Annalena Düker, Das Management von Produktrückrufen: Einflussfaktoren auf die Rückholung von Verbraucherprodukten 460

Advisory Editorial Board:

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

Published by Junior Management Science e.V.

Junior Management Science 4(3) (2019) 305-338

Junior Management Science

journal homepage: www.jums.academy

Portfolio Optimization and Ambiguity Aversion

Belinda Kellerer

Ludwig-Maximilians-Universität München

Abstract

This thesis analyses whether considering ambiguity aversion in portfolio optimization improves the out-of-sample performance of portfolio optimization approaches. Furthermore, it is assessed which role ambiguity aversion plays in improving the portfolio performance, especially compared with the role of estimation errors. This is done by evaluating the out-of-sample performance of the approach of Garlappi, Uppal and Wang for an investor with multiples priors and aversion to ambiguity compared to other portfolio optimization strategies from the literature not taking ambiguity aversion into account. It is shown that considering ambiguity aversion in portfolio optimization can improve the out-of-sample performance compared to the sample based mean-variance model and the Bayes-Stein model. However, the minimum-variance model and the model of naïve diversification, which are both independent of expected returns, outperform the approach considering ambiguity aversion for most of the empirical applications shown in this thesis. These results indicate that ambiguity aversion does play a role in portfolio optimization, however, estimation errors regarding expected returns overshadow the benefits of optimal asset allocation.

Keywords: portfolio choice; asset allocation; estimation error; ambiguity; uncertainty.

1. Introduction

1.1. Problem definition

The classical mean-variance portfolio optimization approach, established by Harry Markowitz, is based on the assumption that expected asset moments are known. In real-life investment decisions, these input parameters are, however, unknown and need to be estimated. Typically, the input parameters are estimated from realized returns and as a result, they are inevitably estimated with error. This is particularly important because the choice of input parameters has a huge effect on the optimal portfolio weights and consequently on the portfolio performance. As these estimation errors are ignored in the classical portfolio optimization approach, a bad out-of-sample performance is the result. Mean-variance optimized portfolios often show a low degree of diversification and extreme portfolio weights[1]. To improve the performance of the classical mean-variance approach, new methods were developed that explicitly take into account that the estimated asset moments are not the true values. In this context, a very important class of methods is the class of Bayesian approaches. Using these approaches, the input parameters are adjusted before they are used to determine the optimal portfolio. This is done by combining a priori information with the historical data by the means of Bayesian updating. Empirical studies have shown that these adjustments lead to a better out-of-sample performance (e.g. Jorion, 1986).

Bayesian approaches, however, assume that the probability distribution of outcomes is known and the decision-maker has a unique prior for the outcomes, while it is ignored that this is not the only possible probability distribution. From the uncertainty about the true probability distribution, a new type of uncertainty arises, called ambiguity. In this thesis, ambiguity is defined as Knightian uncertainty (Knight, 1921), which refers to uncertainty that can't be measured because the probability distribution is unknown and risk refers to measurable uncertainty with a known probability distribution. In 1961, Ellsberg was the first to show experimentally that investors are averse against this type of uncertainty (Ellsberg, 1961). He showed that the majority of investors prefer a risky investment to an ambiguous investment. A central Bayesian assumption is that all uncertainties can be reduced to risks. This assumption, however, contradicts Ellsberg's empirical results. If an investor is ambiguity

[1] See for example Hodges and Brealey (1973), Best and Grauer (1991) and Black and Litterman (1992)

DOI: https://doi.org/10.5282/jums/v4i3pp305-338

averse, his behavior in the Ellsberg experiment leads to a violation of Savage's subjective expected utility theory (Savage, 1954), because the sure-thing-principle is disobeyed. Therefore, ambiguity averse behavior is not rational according to subjective expected utility theory. As a consequence, new portfolio optimization models were developed that allow for ambiguity aversion by weakening the sure-thing principle[2].

In the center of this thesis is the multi-prior approach introduced by Garlappi et al. (2007). They developed an approach that takes both, parameter uncertainty about the true asset moments and ambiguity aversion into consideration. They did so by extending the classical mean-variance approach by two main points. Firstly, they implement an additional constraint on the expected return for each asset to lie within a confidence interval of the estimated expected return. Secondly, they apply an additional minimization over the choice of possible expected returns. Ambiguity aversion is implemented by this additional minimization. Furthermore, the model of Garlappi, Uppal and Wang is also able to account for model uncertainty. This is in case the investor forms his beliefs about expected returns from a factor model, like the Capital Asset Pricing Model (CAPM), but is unsure whether this factor model is the true return-generating model. Garlappi, Uppal and Wang present in their empirical study, that accounting for ambiguity aversion, when optimizing portfolios, improves the portfolio performance and leads to a higher out-of-sample Sharpe ratio compared to the classical mean-variance approach and the Bayes-Stein approach.

This thesis focuses on the question whether the out-of-sample performance can be improved if ambiguity aversion is considered in portfolio optimization, and which role ambiguity plays in improving the portfolio performance. To answer these questions, an empirical study is performed to compare the performance of the ambiguity-averse approach introduced by Garlappi, Uppal and Wang with the performance of several other models from the literature, like the classical mean-variance model, the minimum-variance model, the Bayes-Stein model and the model of naïve diversification. Additionally, the results are tested for robustness by varying important model assumptions. In this way, it is analyzed which additional value can be generated by taking ambiguity aversion into consideration. It is of particular interest which role ambiguity aversion plays compared with the role of estimation errors. Furthermore, the relationship between risk aversion and ambiguity aversion is analyzed and its role for the optimal portfolio is discussed. Besides investigating the effect of ambiguity aversion on the out-of-sample performance of optimized portfolios, a critical view is taken on the topic. Crucial issues, like the rationality of ambiguity averse investment decisions and learning under ambiguity, are discussed.

[2]See for example Schmeidler (1989) and Gilboa and Schmeidler (1989)

1.2. Method of investigation

To analyze the effect of ambiguity averse investment decisions on the portfolio performance, an empirical study is performed. This empirical study is based on the work of Garlappi, Uppal and Wang and is divided into three parts. In the first part, the original results of Garlappi, Uppal and Wang are replicated with a special focus on differences in results. With regard to this empirical application, the investor can build his portfolio from eight international equity indices. In the second part, important input parameters like the sample size to determine the optimal portfolio weights, the degree of risk aversion and the timespan are varied. Finally, in the third part of the empirical study, the investment universe is changed and the portfolio optimization problem is applied to the German DAX30 stocks. Most assumptions are thereby taken from Garlappi, Uppal and Wang to ensure the comparability of the results. This empirical design is chosen to show whether incorporating ambiguity aversion to the portfolio optimization problem leads to a better out-of-sample performance independent of the framework conditions. Furthermore, the variation of input parameters ensures that the improved out-of-sample performance results from taking ambiguity aversion into account and not from other characteristics, for example time-specific or asset-specific characteristics.

The rest of this thesis is organized as follows. In Chapter 2, the foundations of modern portfolio theory are presented and the problems arising from the classical mean-variance approach are discussed. Subsequently, suggestions for improvement, like imposing additional constraints or using Bayesian approaches, are examined. Chapter 3 gives a critical literature overview regarding ambiguity aversion in the context of portfolio optimization. In Chapter 4, the ambiguity-averse approach introduced by Garlappi, Uppal and Wang, which takes ambiguity aversion as well as parameter and model uncertainty into account, is presented. Chapter 5 describes the empirical applications and illustrates the out-of-sample performance of different portfolio optimization strategies in an empirical study by first replicating the original results of Garlappi, Uppal and Wang and then changing important framework conditions. The conclusion is presented in Chapter 6. Further illustrations are collected in the Appendix.

2. Portfolio choice

2.1. Modern portfolio theory

More than 60 years ago, Harry Markowitz established a whole new concept of portfolio optimization by focusing on a holistic approach of several assets to build a portfolio rather than to restrict the investment to a single seemingly profitable asset. In his article "Portfolio selection" of 1952, he derived a normative decision rule to build efficient portfolios. With this decision rule, he laid the foundation for a theory, which was later called modern portfolio theory. About 40

years later, in 1990, Harry Markowitz won the Nobel prize together with Merton H. Miller and William F. Sharpe for their pioneering work in the theory of financial economics[3].

According to Markowitz an investor should focus on two main elements when selecting the optimal portfolio, namely portfolio return and portfolio risk, where risk is defined as fluctuations in returns. He argues that each investor has different interests when choosing a portfolio, but every investor pursues the goal to achieve high returns at low risk. "It seemed obvious that investors are concerned with risk and return, and that these should be measured for the portfolio as a whole" (Markowitz, 1991, p. 470). Modern portfolio theory is based on the assumption that investors are only willing to hold a risky asset if they are compensated with a risk premium in addition to the return on the risk-free asset. Therefore, Markowitz implicitly assumes that investors are risk averse.

To determine the portfolio risk, it is important to analyze the correlations between the assets. If assets are not perfectly correlated with each other, it is possible to reduce the portfolio risk through diversification, without reducing the expected return. As a result, the portfolio risk is always below the weighted average of the individual assets' risk. As the number of assets in the portfolio increases, the role of the covariances becomes more important compared to the role of the variances of the single assets. The direct implication of this concept for the optimal portfolio is that "it is necessary to avoid investing in securities with high covariances among themselves" (Markowitz, 1952, p. 89).

Modern portfolio theory is based on numerous assumptions about the investor and the market. Besides the key assumption that investors seek to maximize yields while minimizing risk, Markowitz assumes that markets are efficient, that investors are rational and expected utility maximizers according to Bernoulli and that they make their decisions based on an individual risk function. Furthermore, the planning horizon in modern portfolio theory covers only one period. At the end of each period the investor has to make a new decision about the allocation of his capital. Asset allocation according to Markowitz is therefore a static decision with a short-term character.

In his article of 1952, Markowitz distinguishes two major steps to determine the optimal portfolio. The first step is to form expectations about future asset moments and the second step is to specify the optimal portfolio. His model is limited to the second step, which can further be divided into identifying the set of efficient portfolios and identifying the investor-specific optimal portfolio using the concept of Bernoulli expected utility maximization. From the expectations about future asset moments, the set of efficient portfolios is built. According to modern portfolio theory, a portfolio is efficient if there is no other portfolio with a higher expected return given a specific level of risk or with a lower risk given a specific level of expected return. All portfolios that are efficient lie on the efficient frontier, which represents the relationship between risk and return. In an efficient portfolio, assets with high expected returns, small variances and low correlation coefficients among each other are the ones with the highest weights. After determining the set of efficient portfolios, the next step is to select the investor-specific optimal portfolio depending on his degree of risk aversion. The optimal portfolio can be derived as the point of tangency between the efficient frontier and the highest possible Bernoulli utility curve. According to Markowitz, the optimization problem given N risky assets can be written as:

$$\max_{w} w^T \mu - \frac{\gamma}{2} w^T \Sigma w \tag{1}$$

in which w is the vector of portfolio weights, μ is the vector of returns, γ is the investor-specific degree of risk aversion and Σ is the variance-covariance matrix.

Markowitz, however, only analyzed how to identify the optimal portfolio given the information about future asset moments, but not how these input parameters can be determined. Since information about future asset moments is unknown, the input parameters need to be estimated. Expected asset moments are typically estimated from realized returns and therefore, they are inevitably estimated with error. As opposed to that, in modern portfolio theory it is assumed that expected returns are known and as a consequence, statistical estimations are assumed to be the true values (Kalymon, 1971). Hodges and Brealey (1973) and Jorion (1985), however, show in empirical studies that historical returns only have a small forecasting power for future returns. Additionally, Kempf et al. (2002) expose in their simulation study that the estimated expected returns differ significantly from the "true" values. They also show that there is a huge gap between the optimal portfolio weights using the "true" parameters and using their estimations from historical returns. Therefore, they argue, that a rational investor should take uncertainty about estimated asset moments into account when selecting a portfolio. Barry (1974) shows that if uncertainty about future asset moments is integrated to the model, the portfolio risk can be reduced while leaving the portfolio return unchanged, leading to increasingly efficient portfolios. Chopra et al. (1993) argue that the estimation of variances and covariances can be performed more precisely than the estimation of expected returns but at the same time, estimation errors in returns have a much higher quantitative influence on the portfolio weights and the resulting portfolio performance compared to estimation errors regarding variances and covariances. Therefore, they conclude that the focus of the estimation should lie on expected returns.

One of the main drawbacks of modern portfolio theory is, that the resulting optimal portfolios are often characterized by instable portfolio weights and by extreme, non-intuitive positions (Black and Litterman, 1992). A small increase in the expected return of only one asset already leads to the reallocation of half of all assets in the portfolio, while the portfolio risk and return hardly change (Best and Grauer, 1991).

[3]For further information see Nobelprize.org

This result is not consistent with the investors' preference to allocate capital to stable portfolios that don't require frequent reallocations, since reallocations are associated with high transaction costs. Furthermore, using mean-variance optimization, the capital is allocated only to a few assets, leading to undiversified portfolios (Black and Litterman, 1992, Green and Hollifield, 1992, Broadie, 1993). These extreme portfolio weights conflict with the desire of diversification. Finally, Jobson and Korkie (1980) and DeMiguel et al. (2009) show that the optimal portfolio selected, using mean-variance optimization, can be even beaten by a uniformly diversified portfolio regarding the Sharpe ratio.

Michaud (1989) argues that the classical mean-variance approach puts high weights on assets with high expected returns, small variances and low covariances among each other, but at the same time, assets with these characteristics are often afflicted with the highest estimation errors. Therefore, assets with highly overestimated returns and with highly underestimated variances are overweighted in the optimal portfolio. This leads to an increase of the estimation error through optimization. For this reason, Michaud among others states that mean-variance portfolio optimization is "error maximization". To solve this problem, Michaud suggests an alternative, statistical, approach to estimate future asset moments. Michaud argues that the observable historical data represents only one realization of the data-generating process. Therefore, he proposes to reestimate the return distribution using Monte Carlo Simulation. From each simulated return distribution, a new set of input parameters for the subsequent optimization arises, leading to resampled efficient portfolios. Scherer (2003), however, criticizes the approach suggested by Michaud because it does not solve the cause of the bad out-of-sample performance of mean-variance portfolios induced by estimation errors regarding expected asset moments. Each portfolio, which is constructed during the simulation, is derived using the same input data. As a result, every resampled efficient portfolio has a similar deviation from the "true optimal portfolio" (Scherer, 2003). Furthermore, Fletcher and Hillier (2001) compared the performance of resampled efficient portfolios with the performance of classical mean-variance portfolios and they don't find any outperformance of either strategy.

Since the bad out-of-sample performance of mean-variance portfolios arises from estimation errors associated with expected asset moments (Merton, 1980), alternative approaches that are explicitly designed to reduce estimation errors, were established. In the following Chapters, different suggestions for improvement, like setting up additional constraints or adjusting the input parameters by the means of Bayesian approaches are analyzed and compared.

2.2. Implementation of additional constraints

Since different out-of-sample studies have shown that portfolios that are optimized using the classical mean-variance approach tend to be very undiversified, additional constraints can be introduced to force a higher diversification. Examples of constraints include the restriction of short selling, upper bound constraints on the position in a single asset or upper bound constraints on the exposure given to a certain country or industry (Brandt, 2009). Often, these restrictions are necessary because of legal requirements and therefore, constraints are a realistic assumption. Introducing additional constraints limits the set of possible portfolio combinations, leading to an efficient frontier that always lies below the non-restricted efficient frontier. As a result, adding constraints can never improve the ex-ante portfolio performance (Grauer and Shen, 2000). However, it has been shown that the out-of-sample portfolio performance can be improved. Restricting portfolios has a smoothening effect on portfolio weights, leading to a higher degree of diversification and to less extreme asset positions (Grauer and Shen, 2000). In a simulation study, Frost and Savarino (1988) show that introducing short selling constraints and upper bound constraints on assets improves the portfolio performance. They measure performance using certainty equivalent returns, which is the difference in portfolio utilities, if the portfolio weights are based on estimated returns and if they are based on the true parameters. Frost and Savarino conclude that the implementation of constraints is reasonable if estimated asset moments are afflicted with estimation errors. This is because the use of constraints prevents the inappropriate emphasis of estimation errors in portfolio optimization (Chopra and Ziemba, 1993).

If estimation errors, however, are not the cause of badly diversified portfolios, constraints are counterproductive, since they predetermine the optimal portfolio to a large extent ex-ante. This fact is criticized by Banz, as he states that constraints are implemented "to force the results which one desires" (Banz, 1997, p. 398). Clasing and Rudd (1988) also argue that implementing constraints is counterproductive. They state that adding restrictions implies that the investor doesn't trust the input parameters. Therefore, they conclude that it is more intuitive to directly adjust the input parameters instead of implicitly changing them by implementing constraints. Grinold and Easton (1998) agree and argue that implementing constraints doesn't treat the cause of the bad out-of-sample performance of mean-variance optimized portfolios.

Out-of-sample studies have shown that the portfolio performance can be improved by implementing constraints, but the performance is still far from good. The reason for this result is that estimated asset moments are afflicted with estimation errors that are so high that the adjustment of the optimal portfolio by adding constraints is not sufficient to solve this problem (Brandt, 2009). To overcome the counterintuitive approach of restricting the outcome by implementing constraints, new techniques have been introduced that directly adjust the input parameters. One important class of these approaches is the class of Bayesian approaches, presented in the next Chapter.

2.3. Bayesian approaches

2.3.1. statistics

The founder of Bayesian statistics is Thomas Bayes (1702-1761). The Bayes' theorem, that was developed by him, is a fundamental theorem in the theory of probability. Bayes' rule describes the probability of an event based on prior knowledge, where $P(A)$ is the prior, the initial degree of belief in A and $P(A|B)$ is the posterior, the degree of belief in A after accounting for B. Bayes' rule can be written as follows:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \tag{2}$$

In the course of portfolio optimization, the basic idea of using Bayes' rule is, to combine a priori available information with the data to determine the posterior distribution of returns. These returns are then subsequently used for portfolio optimization. A priori information may result from financial research, news, events, macroeconomic analysis or asset pricing theories (Avramov and Zhou, 2010). A basic assumption when using Bayesian approaches in portfolio optimization is that returns are independent and identically distributed variables[4]. In the context of portfolio optimization, Bayes' rule can be written as follows (Gelman et al., 1995):

$$P(\theta|z) = \frac{P(z|\theta)P(\theta)}{\int P(\theta)P(z|\theta)d\theta} = \frac{P(z|\theta)P(\theta)}{P(z)} \tag{3}$$

in which $z = (z_1, z_2, z_3, \ldots, z_t)$ is a vector of realized returns and θ represents the unknown parameters of the probability distribution. The value of θ shall be estimated by the means of Bayesian statistics. Bayesian statistics considers θ a random variable. Information that is known ex ante about θ is combined in the a priori probability distribution $P(\theta)$. This a priori distribution is then combined with the information from the historical sample to form the a posteriori distribution $P(\theta|z)$. In contrast to methods of classical statistics where it is assumed that $\theta = th\hat{e}ta$ and therefore a point estimate of θ is provided and any potentially relevant prior information is disregarded, Bayesian statistics provides a whole distribution of θ. As a consequence, it is possible to directly quantify the uncertainty about θ (Gelman et al., 1995). To solve portfolio optimization problems, it is necessary to estimate future asset returns z. This is done by the forecasting distribution $P(\hat{z}|z)$, in which $\hat{z}$ is the estimated value of z. The forecasting distribution results from the likelihood function of a future observation $P(\hat{z}|\theta)$ and the a posteriori distribution $P(\theta|z)$. Using Bayesian statistics, the distribution of future expected returns is derived by integrating over all possible values of θ weighted with their respective a posteriori probabilities (Gelman et al., 1995):

$$P(\hat{z}|z) = \int P(\hat{z}, \theta|z)d\theta = \int P(\hat{z}|\theta)P(\theta|z)d\theta \tag{4}$$

[4]See for example Jorion (1986), Black and Litterman (1992) and Pástor (2000).

Already during the 1970s, the idea to use Bayesian statistics to account for estimation errors in portfolio optimization was developed by Barry (1974) and Klein and Bawa (1976). Unlike implementing constraints, Bayesian approaches exhibit a decision-theoretical fundament. If a Bayesian approach is applied, the adjustment of the optimal portfolio takes place by directly changing the input parameters. In this way, the problem of estimation errors is directly addressed. The adjusted input parameters are then used for portfolio optimization. Barry and Klein and Bawa assume a non-informative diffuse a priori distribution, which means that they assume that all values for the future returns are considered equal probable a priori. As estimation errors are integrated in the Bayesian framework, assets are riskier since parameter uncertainty is an additional source of risk (Avramov and Zhou, 2010). As a result, in the studies of Barry and Klein and Bawa, the vector of expected returns remains unchanged but the variance-covariance matrix is multiplied by $1 + 1/n$, in which n is the sample size. Consequently, if a non-informative prior is used, the set of efficient portfolios doesn't change, but the investor chooses a portfolio further to the left on the efficient frontier compared to the optimal portfolio under the classical mean-variance approach.

Brown (1976) concludes that using Bayesian approaches leads to quantifiable different portfolio structures. He adds that the results, however, depend strongly on the choice of the prior. If a non-informative prior is used, as done by Barry and Klein and Bawa, the only difference to the mean-variance approach is the multiplication of the variance-covariance matrix with the factor $1 + 1/n$ and therefore, the effect on the optimal portfolio is very small if the sample size n is large. If n is very large, the effect is so small that the optimal portfolio is almost identical to the optimal portfolio under the classical mean-variance approach. Avramov and Zhou (2010) state that, for all practical sample sizes, the effect of diffuse priors is negligible and therefore, in order to exploit the decisive advantage of the Bayesian approaches, it is necessary to gain informative priors for example from events, macro conditions or asset pricing theories.

2.3.2. Shrinkage estimators

Shrinkage estimators go back to Charles Stein et al. (1956). Stein points out that the sample mean is not an optimal estimator for the mean of a multivariate normal distributed random variable. He argues, that if expected asset returns are estimated from the sample mean, the value is based solely on the return history of the particular asset and other potential information from the return histories of other assets are disregarded. To overcome this weakness, Stein developed an estimator that includes the return histories of all assets to determine expected returns, leading to more precise results. Stein shrinks the sample mean of each asset towards the overall sample mean of all assets, the "grand mean". This approach has a smoothening effect and prevents extreme values because expected returns of assets with comparably high or low past returns are strongest shrunk towards the grand mean. The key element of the Stein estimator is the shrink-

age factor, which determines how strong the sample means are shrunk towards the grand mean. The general form of the Stein estimator (James and Stein, 1961) can be written as follows:

$$\bar{r}_j(\phi) = \phi r_0 + (1-\phi)\bar{r}_j \tag{5}$$

in which ϕ is the shrinkage factor, r_0 is the shrinkage target, $\bar{r}_j$ is the sample mean of asset j and $\bar{r}_j(\phi$ is the adjusted weighted mean of asset j. The shrinkage factor increases in the number of assets N, decreases in the length of the sample size n and decreases in the distance between the sample mean and the shrinkage target. Shrinkage estimators are, for example, used by Jobson (1979), Jorion (1986), Frost and Savarino (1986) and Chopra et al. (1993). Chopra, Hensel and Turner suggest a global index as shrinkage target and show that a weighted average of the individual asset mean of each country and the global index is a superior estimator of the true asset mean for each country. Frost and Savarino specify an informative prior and assume that all assets have a priori identical expected means and variances and that the pairwise correlation coefficient between any two assets is the same. They shrink the estimates of each assets' expected return, variance and pairwise correlation coefficient, determined from the historical sample, towards the average return, average variance and average correlation coefficient of all assets within the population. Therefore, the approach leads to a shrinkage towards the equal weighted portfolio and the degree of shrinkage depends on the degree to which the sample is consistent with the shrinkage target. Frost and Savarino argue that parameters with large discrepancies between the sample estimate and the overall mean of that parameter are more likely to contain estimation errors and those values are adjusted strongest towards the grand mean, reducing the estimation error. The authors show that the portfolio performance can be improved by applying their informative prior compared to applying an uninformative prior or the classical mean-variance approach.

Jagannathan and Ma (2003) point out that certain constraints that were discussed in Chapter 2.2 can be interpreted in a similar way as shrinkage estimators. They show that no short sales constraints are equivalent to reducing the estimated covariances and that upper bound constraints are equivalent to increasing the respective estimated covariances. Although the effect of both methods can be very similar, the motivation is different. While imposing constraints restricts the outcome exogenously, the adjustment of the input parameters when using shrinkage estimators is endogenous.

2.3.3. Bayes-Stein estimator

Jorion (1985) introduced the Bayes-Stein estimator, a further development of the classical Stein estimator. In contrast to Frost and Savarino, he assumes that estimation errors regarding the variance-covariance matrix are negligible and that variances can be estimated directly from the historical sample. Jorion argues that the minimum-variance portfolio is the only portfolio on the efficient frontier that is free of estimation errors since the determination of the minimum-variance portfolio depends solely on the variance-covariance matrix, which is assumed to be known. The more the investor moves to the right on the efficient frontier, the larger is the influence of estimation errors regarding expected returns. Therefore, the idea of Jorion is to shrink the mean-variance portfolio towards the minimum-variance portfolio, since it is robust to uncertainty about expected means and at the same time it is associated with the least risk. Using the Bayes-Stein shrinkage estimator, the sample means of the individual assets are shrunk towards the mean of the minimum-variance portfolio. The intensity of the shrinkage depends on the sample size n and on the precision of the sample mean v. The shrinkage factor ϕ can be computed as follows (Jorion, 1986):

$$\phi = \frac{v}{n+v} \tag{6}$$

$$\text{with } v = \frac{N+2}{(\bar{r}-r_0\mathbf{1})^T\Sigma^{-1}(\bar{r}-r_0\mathbf{1})} \tag{7}$$

in which N is the number of assets, r_0 is the shrinkage target and $\bar{r}$ is the sample mean. The shrinkage factor ϕ is then inserted into (5) to obtain the weighted average mean for the individual asset. Since the values for ϕ and r_0 are determined directly from the sample, this approach is also called an empirical Bayes approach.

If the Bayes-Stein estimator is used to determine the optimal portfolio, the efficient frontier and the portfolio weights of the efficient portfolios don't change, but the investor will choose a portfolio more to the left on the efficient frontier, closer to the minimum-variance portfolio (Jorion, 1986). Jorion concludes that using the Bayes-Stein shrinkage estimator leads to a significantly improved out-of-sample performance. Chopra et al. (1993) confirm this result.

2.4. Model Uncertainty

The approaches, that have been introduced in Chapter 2.3.2 and 2.3.3, dealt with the adjustment of the sample asset moments by shrinking them towards a grand mean, by the means of Bayesian statistics. All approaches presented in the previous Chapters have in common, that solely the historical sample is used as input data to determine expected asset moments. Findings from asset pricing theory haven't been taken into account so far. In this Chapter, an alternative approach to estimate expected asset moments based on factor models is presented. Chan et al. (1999) show that using factor models to determine the input parameters for portfolio optimization significantly improves the performance of the resulting portfolios. They add, that the resulting portfolios also show a higher degree of diversification compared to a solely data-driven approach.

Pástor (2000) and Pástor and Stambaugh (2000) argue that if expected asset moments are determined exclusively from historical returns, potential information from asset-pricing models are ignored. They developed the data and model approach, in which they combine information from historical data with implications from the CAPM by the means of Bayesian statistics. They define a distinction of cases, with $\omega = 0$ meaning the investor does not believe in the return-generating model and $\omega = 1$ meaning the investor believes dogmatically in the CAPM, where ω can take values between 0 and 1. In the data and model approach, the sample means are shrunk towards the market portfolio and the degree of shrinkage depends on the shrinkage factor. The shrinkage factor measures the importance that is assigned to the CAPM and depends on the investors beliefs in the validity of the CAPM ω and on how well the historical returns can be explained by the CAPM (Pástor, 2000). Pastor and Pastor and Stambaugh argue that the investor's uncertainty about the models pricing ability can be represented by an informative a priori distribution of Jensen's alpha. If the a priori distribution of Jensen's alpha is located around zero, there is only a low degree of uncertainty and the optimal portfolio will be close to the market portfolio. If the a priori distribution of Jensen's alpha, however, is diffuse, the investor is highly uncertain about the models pricing ability and the optimal portfolio is mainly based on realized returns.

When comparing different factor models, Chan et al. (1999) show that no clear favorite specification emerges. A simple single factor model like the CAPM can't be essentially outperformed by a complex multi-dimensional k-factor model. Although the CAPM performs quite well in the study of Chen et. al., it can't capture all of the covariation among assets which could result in a systematically biased estimate of asset moments (Brandt, 2009). Consequently, the difficulty of implementing a factor model in practice is the choice of factors, since there are huge disagreements among researchers about the predictability of returns. Financial economists have identified several economic variables that predict future asset returns. However, there are large disagreements about the "true" predictive regression specification. These disagreements lead to a major source of uncertainty – model uncertainty (Avramov and Zhou, 2010). If the estimation of asset moments is based on assumptions that don't correspond with reality, estimation errors occur because of model uncertainty. The focus on one single model fails to include model uncertainty and therefore the overall uncertainty is underestimated (Cremers, 2002). If posterior odds ratios are used to rank several models and only the best model is used, it is implicitly assumed that the chosen model is true with a unit probability and model uncertainty isn't considered. Subsequently, insights from all other models are completely ignored resulting in a loss of important information (Brandt, 2009). Pástor (2000) adds that in general a model will neither be a complete reflection of reality nor will it be completely useless.

In Chapter 2.3, Bayesian approaches were introduced that incorporate parameter uncertainty by generating a weighted average of the data and the prior. This approach will now be extended to incorporate model uncertainty to portfolio optimization by implementing a weighted average of the competing return-generating models. Taking model uncertainty into account means that not only parameters, but also models, are implemented using probability distributions (Hoeting et al., 1999). The Bayesian model averaging approach (BMA), presented by Hoeting et al. (1999) determines posteriori probabilities to a set of competing models and then these probabilities are used as weights for the respective model to form an overall composite model. This overall model is then used to solve portfolio choice problems. The weights depend on the ability of the model to fit the data and on prior beliefs in the model (Hoeting et al., 1999; Raftery et al., 1997). In this context, a Bayesian approach is preferred because it allows to directly incorporate model uncertainty and is robust to model misspecification (Avramov and Zhou, 2010). Model averaging leads to higher estimates of variances compared to approaches that don't account for model uncertainty. This is why ignoring model uncertainty leads to overconfident decisions (Nigmatullin, 2003). Avramov (2002) and Cremers (2002) each introduce a Bayesian model averaging approach which integrates out both, parameter and model uncertainty. In this context, Avramov states that investors who ignore model uncertainty suffer from significant utility losses. He shows for six portfolios over the time period 1953 to 1998, that the composite model outperforms the individual model with the highest posterior probability for any model selection criterion used. Cremers stresses the high importance of taking model uncertainty into account and shows that BMA provides an improved forecasting ability compared to model selection approaches. Anderson and Cheng (2016) develop a BMA approach in which the investor updates the probabilities of each model after every period using Bayes' rule. They conclude that accounting for model uncertainty improves the out-of-sample portfolio performance.

3. Ambiguity aversion

3.1. Concept of ambiguity aversion

3.1.1. A definition of ambiguity

It is well known that agents are risk averse and are willing to sacrifice expected return in order to reduce risk. But what is meant by risk? Knight (1921) was the first to note that not all sources of uncertainty can be quantified probabilistically and therefore risk is radically different from uncertainty. Knight distinguishes between two different types of uncertainty. Risk is referred to situations in which the probabilities of the outcomes are known, whereas ambiguity is referred to situations in which the probability distribution is unknown. Ambiguity might occur because the decision maker is unable or unwilling to summarize the available information by a unique probability distribution, due to a lack of information.

3.1.2. The Ellsberg-experiment

While Knight was the first to differentiate between risk and ambiguity, Ellsberg (1961) was the first to show in his experiments that decision makers behave ambiguity averse.

In his first two-color urn-experiment there are two urns with 100 balls each. Urn1 contains exactly 50 red balls and 50 black balls while Urn2 also contains 100 balls, but the distribution of red balls and black balls is unknown and all combinations, ranging from 100 red balls and 0 black balls to 0 red balls and 100 black balls, are possible. Urn1 is a risky urn since the probability distribution is known. Urn2, however, is an ambiguous urn because the decision maker is unable to form a unique prior distribution over Urn2 with certainty, due to a lack of information. Decision makers were then asked whether they preferred (i) to bet on a red ball from either Urn1 or Urn2 and whether they preferred (ii) to bet on a black ball from either Urn1 or Urn2. In Ellsberg's experiment, most decision makers preferred a bet on a red ball from Urn1 to a bet on a red ball from Urn2, and a bet on a black ball from Urn1 to a bet on a black ball from Urn2. These results indicate that decision makers act ambiguity averse, in the sense that they prefer situations with known probabilities to situations with unknown probabilities.

In the second three-color urn-experiment, there is one urn that contains 90 balls. There are exactly 30 red balls in the urn and 60 balls that are either black or yellow and the distribution of black and yellow balls is unknown. The decision maker has to choose between a bet on (iii) a red ball and a bet on a black ball and between a bet on (iv) a ball that is either red or yellow and a bet on a ball that is either black or yellow. In this experiment, most decision makers preferred to bet on a red ball in situation (iii) and to bet on a ball that is either black or yellow in situation (iv). This behavior confirms the result that decision makers are ambiguity averse.

While Ellsberg's urn-experiments show the relevance of ambiguity in decision making, ambiguity averse behavior violates axioms that are necessary to derive the decision maker's subjective expected utility (Ellsberg, 1961). An agent who prefers a bet on a red ball from Urn1 to a bet on a red ball from Urn2 (i) indicates that he believes that a red ball from Urn1 is more probable than a red ball from Urn2. If the same agent prefers a bet on a black ball from Urn1 to a bet on a black ball from Urn2 (ii), this choice indicates that he believes that a black ball from Urn1 is more probable than a black ball from Urn2, which corresponds to the belief that a red ball from Urn1 is less probable than a red ball from Urn2. These two beliefs apparently contradict each other. The typical response from the second three-color urn-experiment directly violates Savage's sure-thing principle (Savage, 1954). The sure-thing principle requires that rankings of acts are independent of common parts. In the Ellsberg-experiment, situation (iii) is extended by the yellow ball in situation (iv), which is the common part. Since adding the yellow ball shifts the preference from a bet on the red ball towards a bet on a ball which is either black or yellow, this behavior violates Savage's subjective expected utility theory.

3.1.3. Ambiguity aversion in the stock market

Ellsberg has shown that agents behave ambiguity averse in an experimental urn setting. Bossaerts et al. (2010) show that ambiguity aversion also plays an important role in asset markets. Investing in the stock market involves risk and ambiguity, risk whether the price will rise or fall and ambiguity about the probability distribution of outcomes. Bossaerts et al. (2010) show that ambiguity aversion has an influence on the size and the structure of the optimal portfolio holdings and also on asset prices. They show that if ambiguity is present in competitive markets, an ambiguity averse decision maker wants to hedge against this ambiguity. This desire leads to increasingly diversified portfolios as ambiguity increases. In the study of Bossaerts et. al., a significant fraction of decision makers is sufficiently ambiguity averse so that they even refuse to hold an ambiguous portfolio. Koziol et al. (2011) show that private, but also institutional investors are ambiguity averse and as a result, they reduce their allocations to risky and ambiguous portfolios. In their empirical study, the allocation to ambiguous assets is rather small and significantly below the theoretical optimal allocation. Antoniou et al. (2015) add that the willingness to invest decreases as ambiguity in the stock market increases. In their empirical study, they use dispersion in analysts' forecasts about future returns as a measure of ambiguity and equity fund flows as a measure of market participation. If ambiguity in the stock market increases, market participation strongly decreases. These results indicate that ambiguity aversion doesn't only play a role in an artificial urn environment but also in the everyday investment decisions of private and institutional investors.

3.2. Ambiguity averse behavior under the subjective expected utility theory

Expected utility theory assumes that the probabilities of outcomes are known and that decision makers' preferences can be presented by an individual utility function. The utility can then be computed as the expected utility of the possible outcomes weighted by their probabilities (Von Neumann and Morgenstern, 1945). In contrast, the subjective expected utility theory (SEU) introduced by Savage (1954) recognizes, that probabilities may not be objectively known. In this case, the decision maker forms his own subjective probabilities that are used for decision making. Subjective probabilities go back to Ramsey (1931), who argues that an agent's beliefs can be measured by his willingness to bet. With SEU, Savage introduced the axiomatic derivation of a Bayesian decision problem, in which probabilities and utilities are not objectively given. In the Bayesian paradigm, any source of uncertainty can be quantified probabilistically. This means, that in order to satisfy the set of axioms, even if the probabilities are not objectively known, it is necessary to compute expectations with respect to the return distribution. The Bayesian quantification of uncertainty implies that all uncer-

tainties can be reduced to risks[5] and therefore the decision maker has a unique prior probability. As a result, the investor is only exposed to risk and trades off this risk against expected returns to maximize his expected utility (Aït-Sahali and Brandt, 2001).

Knight and Ellsberg criticize the Bayesian approach and argue that decision makers might not have enough information to form subjective expectations about the probability distribution. In this case, uncertainty can't be quantified probabilistically, which is exactly the definition of Knightian uncertainty. They argue that if investors are confronted with Knightian uncertainty, they might not be able to form a unique prior, leading to ambiguity averse behavior, as has been shown in the Ellsberg-experiment. Furthermore, the Bayesian approach does not differentiate between risk and ambiguity and therefore subjective probabilities are treated in the same way as certain probabilities. Consequently, the Bayesian approach does not represent the confidence the decision maker has in his own probabilistic assessments (Epstein and Wang, 1994). Ultimately, ambiguity averse behavior, as has been shown in the Ellsberg-experiment, cannot be modeled using Savage's Bayesian approach.

One potential path, as argued by Gilboa and Marinacci (2016), is to incorporate ambiguity aversion into the decision-making process by relaxing the assumption that decision makers are Bayesian. They state that a non-Bayesian approach doesn't require the decision maker to form a unique prior probability about expected returns and allows the decision maker to incorporate his uncertainty about the probabilities.

3.3. Decision-making under ambiguity

When decision makers form a belief about the probability distribution of returns, they take information like statistical models or analysis of fundamentals into account. However, the distribution of returns may remain uncertain. In this case the ambiguity averse decision maker will account for this ambiguity in his decisions.

Under the principle of insufficient reason, going back to Jakob Bernoulli, the decision maker assumes that there is no reason to expect one event to be more likely than the other and therefore one should assume that probabilities are distributed equally. Based on this assumption, the alternative is chosen which generates the highest expected utility. Gilboa and Marinacci (2016) argue that in most real-life situations there is too much information available to apply the principle of insufficient reason, but too little information to form a unique prior probability distribution as required by SEU. This dichotomy led to the development of decision rules under ambiguity, for example by taking multiple priors into account.

Many decision rules under ambiguity are based on Wald's rule (Wald, 1950). Wald's rule, which is also called minimax rule, is very conservative, as the decision is based solely on the worst possible outcome while all other outcomes are ignored. Wald's decision rule is based on the decision maker's a priori lack of confidence in his information. Using Wald's rule, the ambiguity averse decision maker assesses all acts by the minimum expected outcome associated with the respective act and then chooses the act leading to the highest minimum expected outcome (Mukerji and Tallon, 2001). Ellsberg (1961) criticizes, that if the decision is based solely on the worst possible outcome, probabilities that are known are completely ignored. He argues that in this way, the application of Wald's rule leads to a distortion of the decision maker's best estimate of probabilities towards less favorable probabilities, in which the degree of distortion depends on the level of confidence the decision maker has in his own estimated probability distribution. The less confident he is, the more he will rely on the worst possible outcome. Nevertheless, Ellsberg states that the minimax criterion might be a good starting point for a conservative decision maker in the presence of high ambiguity. Wald concludes that "a minimax solution seems, in general, to be a reasonable solution of the decision problem when an a priori distribution does not exist or is unknown" (Wald, 1950, p. 18). Applying the minimax rule can lead to a violation of Savage's sure-thing principle and is consistent with the results of the Ellsberg-experiment.

3.4. Incorporating ambiguity aversion into portfolio optimization

3.4.1. Choquet expected utility

The Choquet expected utility model (CEU), introduced by Schmeidler (1989) was the first axiomatically sound non-Bayesian decision model for portfolio choice problems. Schmeidler's starting point was that probabilities should reflect the decision maker's willingness to bet and, under this definition, probabilities don't need to be necessarily additive. Schmeidler defines an expected utility maximization criterion for the nonadditive case, which allows the decision maker's beliefs to be presented by a unique, but not necessarily additive probability distribution. In his model, Schmeidler weakens Savage's sure-thing principle and he models probabilities by a capacity v, a set function that is not necessarily additive. If probabilities are nonadditive, the sum of the probabilities of several mutually exclusive outcomes is not necessarily equal to the probability that one of the outcomes will occur. If the sum of probabilities is smaller than the probability that one of the outcomes will occur, ambiguity aversion is reflected (Dow and da Costa Werlang, 1992). Schmeidler defines a notion of ambiguity aversion, given non-additive probabilities v as follows:

$$v(A) + v(B) \leq v(A \cup B) + v(A \cap B). \tag{8}$$

Schmeidler uses the notion of integration suggested by Choquet (1954) to generalize the case in which the capacity v is additive. Since the capacity is not necessarily additive, the weight of an outcome is determined by its place in the

[5] Ramsey (1931) states that „For a rational man all uncertainties can be reduced to risks".

ranking of all possible outcomes. According to Schmeidler, a decision maker with CEU preferences is ambiguity averse if his capacity v is convex and his utility function is concave or linear.

3.4.2. Maxmin expected utility

The idea of maximizing the minimum expected outcome was formalized in the model of Maxmin expected utility (MEU), introduced by Gilboa and Schmeidler (1989). The model is based on the idea that if there is not enough information available to form a single probability distribution, it is preferable to ask the decision maker to form a whole set of probability distributions P. Gilboa and Schmeidler argue, that if investors are uncertain about the true probability distribution, they shouldn't restrict themselves to one particular distribution as a proxy for the true probability distribution. Furthermore, considering multiple priors is advantageous, since portfolio weights react very sensitive to the choice of a certain prior and investors demand portfolios, that perform good for a whole set of possible probability distributions. The Maxmin expected utility model takes uncertainty about the true probability distribution into account by maximizing the expected utility under the worst-case scenario in the set P. An investor with MEU preferences evaluates each act based on the expected utility given the worst-case probability distribution in the set of prior distributions. However, the model is less conservative than Wald's rule, since not all possible prior distributions, but only a certain choice of priors is considered.

The MEU model is based on an Anscombe-Aumann framework (Anscombe et al., 1963) in which outcomes are modeled as lotteries. Gilboa and Schmeidler extend Savage's axioms by an uncertainty aversion axiom, which is the key axiom behind their model. The uncertainty axiom weakens the independence axiom of Anscombe-Aumann and states that, if the decision maker is indifferent between two uncertain acts f and g, the decision maker's preference can be written as

$$af + (1-\alpha)g \geq f\,(or\ g) \tag{9}$$

in which α is a factor between 0 and 1 (Gilboa and Schmeidler, 1989). The uncertainty axiom states that if the decision maker prefers a combination of two indifferent acts to either of the two acts, the decision maker is ambiguity averse. If the two uncertain acts f and g are combined, a new act arises which is not uncertain. Therefore, the two uncertain acts can be reduced to one risky act and the investor can hedge against the uncertainty by mixing the two acts (Etner et al., 2012). It can be shown that CEU is a particular case of MEU given that the decision maker is ambiguity averse. The CEU model, however, doesn't require the decision maker to be ambiguity averse and is compatible with any capacity.

3.4.3. Literature overview

Following the seminal works of Schmeidler and Gilboa and Schmeidler, many further models incorporating ambiguity aversion, mostly based on CEU and/or MEU, were developed. In the following, important models are presented that give insights on how uncertainty about parameters and/or models can be incorporated to portfolio choice problems. The models introduced in this Chapter show a similar approach as the multi-prior model of Garlappi et al. (2007) presented in Chapter 4, which is used for the empirical analysis in Chapter 5.

Goldfarb and Iyengar (2003) present an optimization model under ambiguity, which is robust to parameter uncertainty. In their model, possible deviations from the expected asset moments, which are estimated from historical data, are modeled as unknown, but bounded. All parameters lie within uncertainty sets that match with confidence regions around the estimate of the parameters. The size of the uncertainty set reflects the confidence level of the decision maker. The optimal portfolio is then determined by assuming worst-case specifications of the parameters.

Tütüncü and Koenig (2004) introduce a robust asset allocation model, in which expected asset moments are modeled as uncertainty sets, which are used to obtain portfolios with the best worst-case behavior. The uncertainty sets correspond to confidence regions and are determined by bootstrapping. For expected returns, the worst-case is defined as the minimum value of the chosen confidence interval, while for the variance, the worst-case is defined as the maximum value of the chosen confidence interval. Tütüncü and König show that using this robust worst-case approach leads to a significantly better out-of-sample portfolio performance and to an increased portfolio stability compared to the mean-variance portfolio. They conclude that their approach is a reasonable alternative for conservative investors.

Wang (2005) presents a minimax approach, in which he determines the optimal portfolio numerically in a Bayesian setting, accounting for parameter and model uncertainty. In his model, uncertainty is depicted as a set of possible priors with different degrees of precision. Wang uses a shrinkage approach, in which the prior probability distribution incorporates data as well as prior beliefs generated from an asset pricing model. The resulting expected mean is a weighted average from insights of the asset pricing model and the data. Wang shows that the optimal portfolio for an investor who is ambiguity averse differs strongly from the market portfolio underlying the CAPM and also from the portfolio based solely on the sample estimate. This result indicates that model uncertainty is as important as parameter uncertainty.

Pflug and Wozabal (2007) also introduce a minimax approach, in which confidence sets are used to determine the probability distribution. The size of the confidence sets depends on the amount of information available. The more data is available, the tighter is the confidence set and the smaller is the resulting cost of ambiguity. Their approach points out the tradeoff between return, risk and model ro-

bustness.[6]

All models, introduced in this Chapter, have in common that they want to formulate optimization problems in a way, that they result in a good portfolio performance for all possible realizations of the unknown parameters. This is why these approaches are also called robust approaches. Tütüncü and Koenig (2004) and Scherer (2007) state that robust portfolio rules create portfolios which react less sensitive to new information and therefore are more robust to changes in expected returns. This is because robust decision rules lead to a shift in the portfolio from assets with high parameter uncertainty and large potential losses associated with parameter uncertainty to assets with less uncertainty, leading to an increased performance (Zhang et al., 2017). The performance of classical mean-variance estimators drops strongly even if the sample distribution differs only slightly from the assumed distribution. The performance of robust estimators reacts much weaker to deviations from the assumed distribution. However, the classical mean-variance estimator outperforms the robust estimator if the underlying distribution is correct (DeMiguel and Nogales, 2009). Chen et al. (2014) show in their empirical study that, if an ambiguity averse decision maker ignores ambiguity, welfare costs can exceed 15% compared to decision-making under robust models.

3.4.4. Smooth models incorporating ambiguity aversion

The models discussed in the previous Chapter have in common that they use a minimax approach and adopt an absolute worst-case. Lim et al. (2012) criticize this approach because only the performance in the worst case is considered, ignoring the concern of underperforming in other, possibly more probable, cases. They argue that adopting an absolute worst-case approach is overly pessimistic and sensitive to the choice of the uncertainty set. In this Chapter, a smooth model incorporating ambiguity aversion is presented which is less conservative than absolute worst-case models.

Klibanoff et al. (2005) were the first to introduce a smooth model incorporating ambiguity aversion[7]. The main characteristic of their model is the separation between ambiguity, which represents the decision maker's subjective beliefs and the ambiguity attitude, which represents the decision makers tastes[8]. As a consequence of this separation, the decision maker doesn't have to select the act that maximizes the minimum expected utility, but is allowed to be less pessimistic. Klibanoff, Marinacci and Mukerji assume that the decision maker has a subjective prior probability over the set of possible probability distributions. The decision maker then computes a certainty equivalent with regard to expected utility for each probability distribution. Ultimately, the decision maker maximizes the expected utility of the conditional certainty equivalents with regard to a second utility function, which indicates the ambiguity attitude of the decision maker. Therefore, the focus is shifted from the minimum expected utility towards an aggregation of all possible expected utilities. The more ambiguity averse the decision maker is, the higher are the weights he attaches to the less favorable probability distributions. Thimme and Völkert (2015) show in their empirical study that the smooth model of Klibanoff, Marinacci and Mukerji improves the fit of the data.

The main difference to the MEU model is, that the decision maker considers various probability distributions, while Gilboa and Schmeidler only consider the worst-case distribution. Consequently, MEU is an extreme case of the smooth model when extreme ambiguity aversion is assumed (Klibanoff et al., 2005).

3.5. Ambiguity aversion and stock market puzzles

3.5.1. Limited market participation

Ambiguity aversion plays a key role in explaining some of the most important asset pricing puzzles, like limited market participation, the equity premium puzzle or the size effect. Epstein and Schneider (2008) and Pataracchia (2011) state that ambiguity aversion affects the way how investors process information. They argue that if the quality of information is not known, investors react ambiguity averse and take a worst-case assessment of quality. In case of good news, the worst-case is that the information is noisy whereas in case of bad news, the worst-case is that the information is precise. This asymmetric reaction to information leads to a shift in mass probabilities from favorable outcomes towards less favorable outcomes (Pataracchia, 2011). In this context, several authors have shown that ambiguity aversion has an effect on market participation and can lead to a situation in which investors find it rational not to participate in the stock market at all.

Dow and da Costa Werlang (1992) state that, for a decision maker with CEU preferences, there exists a price interval of non-zero length, for which the decision maker prefers neither to buy nor to sell the risky asset and instead he strictly prefers to hold a zero position in the risky asset. They argue that this is because the decision maker acts as if he assigns different probability distributions to the long and the short position in the asset. This can lead to a situation in which the decision maker's minimum asking price for a short position in the asset, is higher than his maximum bid price for a long position in the asset. If the asset price lies somewhere in between, the decision maker does not trade at all. This behavior strongly contradicts the SEU framework, in which the decision maker buys the risky asset if the price is below his certainty equivalent and he sells the risky asset if the price is above his certainty equivalent. Under the SEU framework there is no no-trade interval. Cao et al. (2005) and Bossaerts et al. (2010) add that ambiguity aversion among investors is heterogeneous and, as a result for every ambiguity averse

[6] Further models that extend the models of Schmeidler and Gilboa and Schmeidler include Chen and Epstein (2002), Uppal and Wang (2003), Deng et al. (2005), Maccheroni et al. (2006), Ma et al. (2008) and Qu (2015). All these models have in common, that they include CEU and/or MEU as special cases.

[7] The model has been extended in Klibanoff et al. (2009) to a setting involving dynamical decision making.

[8] Ambiguity aversion is investor-specific and doesn't change over time, while ambiguity is asset-specific and does change over time.

investor, there exists an individual price region, for which he prefers not to participate in the market. The size of this region depends on the investor-specific degree of ambiguity aversion. This mechanism explains why ambiguity aversion can lead to limited stock market participation.

Epstein and Miao (2003) use this mechanism, triggered by ambiguity aversion, to explain the home bias, the limited participation in foreign markets compared to the home market. They argue that investors are more certain about the stock returns of companies in their home country and therefore they prefer to invest in the home market.

3.5.2. Equity premium puzzle

The equity premium puzzle describes the disproportionate high premium for risky assets compared to the return on the risk-free asset. Although it is intuitive that there exists an equity premium because of different degrees of risk, the high amount of the equity premium is not consistent with theoretical predictions. However, most theoretical predictions take only risk into account and ignore ambiguity. Cao et al. (2005) argue that the total equity premium can be decomposed into a risk premium and an ambiguity premium. Accordingly, there exists a separate premium for bearing ambiguity additional to the premium for taking risks, explaining the high amount of the total equity premium. Izhakian (2012) defines the ambiguity premium as the amount the decision maker is willing to pay in order to replace a bet with unknown probabilities by a bet with known probabilities, given the same expected outcomes for the ambiguous and the risky bet. Pataracchia (2011) shows that asset prices decrease as ambiguity increases, because investors demand a higher ambiguity premium in order to prefer to hold the ambiguous portfolio. Anderson et al. (2009) show that excess market returns are strongly associated with ambiguity, while they are less strongly associated with risk. They conclude from their findings that the risk premium is dominated by the ambiguity premium. This result is confirmed by the empirical study of Thimme and Völkert (2015), who show that the risk premium contributes to the total equity premium by 31%, while the ambiguity premium contributes to the overall equity premium by 69%.

3.5.3. Size effect

Antoniou et al. (2014) use ambiguity aversion to explain the size effect, a market anomaly, which describes that companies with a small market capitalization outperform companies with larger market capitalizations. The authors argue that analysts' forecasts about expected earnings of smaller companies are more ambiguous, since analysts find it harder to estimate the accuracy of forecasts for smaller companies, due to a lack of information. This leads to a large dispersion in analysts' forecasts about future returns. Due to the high degree of ambiguity in earnings forecasts, ambiguity averse investors will react pessimistically and overweight the worst-case scenario. Therefore, the authors show that the stock prices of small companies are reduced, compared to a situation without ambiguity aversion. At the end of the quarter, the pessimism is resolved as more information gets available and consequently ambiguity regarding expected returns decreases. This reduction of ambiguity then results in increasing stock prices. Antoniou, Galariotis and Read show empirically that this behavior leads to the size effect around earnings announcements.[9]

3.6. Is ambiguity aversion rational?

In Chapter 3.4, models were presented that incorporate ambiguity aversion into portfolio optimization and, in Chapter 3.5, ambiguity aversion was used to explain some of the most important puzzles in finance, indicating that ambiguity aversion does play an important role in real life investment decisions. But even if investors behave ambiguity averse, does that also mean that it is rational?

Raiffa (1961) was the first to argue that a rational decision maker should not be ambiguity averse. Scherer (2007) agrees that investors who are ambiguity averse behave irrational. He especially criticizes the MEU approach introduced by Gilboa and Schmeidler, because their model violates Savage's sure-thing principle and the model can lead to a Dutch Book outcome, a situation where the decision maker agrees to a combination of bets that guarantees him to lose. Sims (2001) agrees that violating the sure-thing principle on purpose shouldn't be recommended to decision makers. Al-Najjar and Weinstein (2009) add that admitting ambiguity averse behavior as rational can not only lead to a Dutch Book outcome, but also to situations in which the decision maker reacts sensitive to irrelevant sunk costs or is averse to information, situations that are clearly irrational. Najjar and Weinstein further criticize that in the ambiguity aversion literature the sure-thing principle is relaxed, while other aspects of Savage's framework are kept. They argue that, if ambiguity aversion is not rational but a behavioral anomaly, it doesn't seem to make sense to assume a rational decision maker in any other aspect of the model. Scherer (2007) concludes, that these drawbacks of MEU are more severe than the Ellsberg paradoxon is for subjective expected utility theory.

Al-Najjar and Weinstein (2009) give an alternative explanation for Ellsberg choices, in which the decision maker believes that the odds of the experiment are adversarial manipulated and therefore he prefers the perceived safer, risky option[10]. They argue that this behavior may be rational in real life situations, in which the counterpart might have superior knowledge about the odds, but in an experimental setting, like the Ellsberg-experiment, ambiguity averse choices are irrational and might arise from a misapplication of real life situations to the experimental setting.

[9] While Antoniou, Galariotis and Read show empirically how ambiguity aversion can create the size effect, Epstein and Schneider (2008) show theoretically, that prices under conditions with a high level of ambiguity are lower in the beginning of the quarter because investors are ambiguity averse. This pessimism about expected returns is then corrected leading to increasing prices, as soon as ambiguity starts to resolve as the quarter comes to a close.

[10] This alternative explanation of the Ellsberg paradoxon is also known as deceit aversion mechanism

A different alternative explanation for Ellsberg choices is given by Kovarik, Levin and Wang (2016), based on complexity aversion. They introduce a model in which complexity aversion, and not ambiguity aversion, is the reason why decision makers prefer the risky urn to the ambiguous urn in the Ellsberg-experiment. In their experiment, they relate complexity to the compoundness and the number of alternatives of the lotteries. They assume that decision makers want to avoid the cognitive effort related to the reduction of compound lotteries and therefore they estimate the probabilities and afterwards they discount the final value. In this way, the decision maker considers the ambiguous urn to be more complex than the risky urn. Since decision makers are assumed to be complexity averse, a higher degree of complexity reduces the utility of the lottery. As a result, a complexity averse decision maker prefers the unambiguous, less complex lottery. According to this model, Ellsberg choices don't necessarily contradict SEU, because complexity aversion affects the investor's utility and not his beliefs.

While there is some doubt about the rationality of ambiguity aversion and there exist also alternative explanations for Ellsberg choices like the deceit aversion mechanism or complexity aversion, the major part of the literature agrees that ambiguity aversion should be incorporated into portfolio choice problems. If investors care about ambiguity in the stock market and their utility depends on the degree of ambiguity, the portfolio decision rule should incorporate ambiguity. For an investor who is not ambiguity averse, however, a different, less conservative, model might be more appropriate.

3.7. Learning under ambiguity

Does ambiguity dissolve as the decision maker learns about his environment? If ambiguity occurs because of a lack of information about the true probability distribution, investors can learn about the missing information and ambiguity will vanish in the long run. Investors are uncertain about many parameters in financial markets and can learn about these parameters by observing data. Although huge amounts of financial data are available, the ability to learn might be limited due to the huge amount of randomness in financial markets (Pastor and Veronesi, 2009). Consequently, there might be circumstances that make it impossible for the decision maker to fully learn about the true probability distribution.

Marinacci (2002) shows that if there is repeated sampling with replacement from the same ambiguous Ellsberg urn, the decision maker learns about the true color distribution and ambiguity resolves asymptotically. However, this in only true in the special case of unambiguous signals. Several models of learning under ambiguity were established which show that ambiguity does not need to vanish completely in the long run if signals are ambiguous. Epstein and Schneider (2007) introduce a model of learning under ambiguity, which is based on the Maxmin expected utility model introduced by Gilboa and Schmeidler (1989). In their model, they distinguish between noisy signals and ambiguous signals. They argue that the draw from the ambiguous urn is an ambiguous signal, while the draw from the risky urn is a noisy signal. Since ambiguity averse decision makers presume the worst-case, the ambiguous signal is treated as if it was less precise than the noisy signal. Regarding the ambiguous urn, signals about new data are always viewed as ambiguous, also in the long run. Therefore, ambiguity about the true probability distribution may persist forever. Epstein and Schneider conclude that if ambiguous signals are considered, a learning process can be observed and ambiguity might be resolved over time, however, it doesn't vanish completely in the long run because at some point in time, the decision maker can't learn anymore about the true parameters.

Campanale (2011) introduces a model of life-cycle portfolio choice, in which investors have minimax preferences and they assume that there is a whole set of possible return distributions. Campanale shows that part of the ambiguity vanishes over time, as investors learn about the true return distribution by observing realized returns. Consequently, the set of possible return distributions shrinks, but investors are not able to identify the true return distribution with certainty. Therefore, Campanale concludes that ambiguity must also be considered in long-term investment decisions.

Branger et al. (2013) show that investors who ignore ambiguity aversion or who refuse to learn experience a utility loss. However, the loss from not learning is much smaller compared to the utility loss from ignoring ambiguity aversion[11]. The authors conclude that taking ambiguity aversion into account is of first order importance, especially since the utility loss from not learning is even reduced as ambiguity aversion increases. Binmore et al. (2012) and Guidolin and Rinaldi (2013) agree and state that learning only has a moderate effect on investment decisions.

Chapter 3 dealt with the topic of ambiguity aversion in portfolio choices. It was shown, that ambiguity aversion plays an important role in experimental settings but also in the stock market. While there is some doubt about the rationality of ambiguity aversion, and also alternative explanations were offered, the major part of the literature agrees that ambiguity aversion should be incorporated into portfolio optimization problems. Different models taking ambiguity aversion into account were presented, for example the CEU model introduced by Schmeidler, the MEU model introduced by Gilboa and Schmeidler or the smooth model incorporating ambiguity aversion introduced by Klibanoff, Marinacci and Mukerji. It was also shown that ambiguity aversion helps to explain some of the most important puzzles in finance, like limited market participation or the equity premium puzzle. The last section dealt with learning under ambiguity, however, it was argued that learning only has a moderate effect on investment decisions.

[11] Branger, Larsen and Munk show that the utility loss from refusing to learn is below 4% of the initial portfolio value even in case of a long investment horizon, while the utility loss from ignoring ambiguity aversion amounts to more than 50% of the initial wealth for an investment horizon of 20 years.

4. Portfolio Selection with Parameter and Model Uncertainty: A Multi-Prior Approach (Garlappi et al., 2007)

4.1. Theoretical framework

This Chapter deals with the multi-prior approach introduced by Garlappi et al. (2007), that incorporates parameter and model uncertainty and takes ambiguity aversion into account. In Chapter 5, this model is then tested against other models in the literature. The model shows similarities with the models of Goldfarb and Iyengar (2003), Tütüncü and Koenig (2004) and Wang (2005), which were presented in Chapter 3.4.3. One important feature of the model is that it is able to account for parameter uncertainty and model uncertainty. Therefore, the model can be implemented if the investor estimates future returns from historical observations but also if the investor relies on a certain factor model to generate expected returns and he is uncertain whether this is the true return-generating model. Another important feature is that the model is based on a solid axiomatic foundation, since it adopts the multi-prior approach (MEU) introduced by Gilboa and Schmeidler (1989).

However, the model takes only estimation errors regarding expected returns into account, while the historical variance-covariance matrix is not adjusted. Therefore, it is assumed that $\Sigma = \hat{\Sigma}$. Although, as mentioned in Chapter 2.1, estimation errors regarding returns have a larger impact on the optimal portfolio weights than estimation errors regarding variances, it is still important to estimate variances precisely, especially because the number of covariances is in a quadratic proportion to the number of estimated returns (Ledoit and Wolf, 2003).[12] The model also does explicitly not consider learning. While there exists some evidence that ambiguity changes over time through learning, as argued in Chapter 3.7, most researchers agree that learning only has a moderate effect.

Starting from the classical mean-variance optimization model, Garlappi, Uppal and Wang (GUW) made two major adjustments to (1). Firstly, they implement an additional constraint on the expected return for each asset to lie within a confidence interval of the estimated expected return. The larger the confidence interval is, the less does the investor rely on the estimated return and as a consequence, the corresponding portfolio weight declines. If the size of the confidence interval equals 0, that is the return is estimated with perfect precision, the portfolio weights are equal to the portfolio weights of the classical mean-variance portfolio. Secondly, to account for ambiguity aversion, they implement a minimax approach, by introducing an additional minimization over the multiple priors of expected returns and/or models. The adjustments to (1) can be written as follows[13]

$$\max_{w} \min_{\hat{\mu}} w^T \hat{\mu} - \frac{\gamma}{2} w^T \Sigma w, \tag{10}$$

$$\text{subject to } f(\hat{\mu}, \Sigma) \leq \epsilon, \tag{11}$$

$$w^T \mathbf{1}_N = 1 \tag{12}$$

with (12) restricting the portfolio weights to sum to 1 and with ϵ being the ambiguity parameter, reflecting both, ambiguity and ambiguity aversion. In this model, no separation between preferences and beliefs regarding ambiguity takes place and the degree of ambiguity aversion is normalized to 1. By varying ϵ, different degrees of ambiguity about the expected returns can be represented. GUW show that if the set of priors is assumed to be Gaussian, the factor ϵ directly presents the size of the confidence intervals. Function (11) depicts the additional constraint and can take different forms, depending on whether ambiguity is estimated asset by asset, jointly for all assets or for subsets of assets[14]. Which form is used depends on whether parameter uncertainty only or both, parameter and model uncertainty are considered.

If one wants to consider parameter uncertainty only, ambiguity about expected returns can be estimated jointly for all assets and the portfolio optimization problem can be solved as follows[15]

$$w_{AA}(\epsilon) = \phi_{AA}(\epsilon) w_{MIN} + [1 - \phi_{AA}(\epsilon)] w_{MV} \tag{13}$$

in which $w_{AA}(\epsilon)$ is the vector of optimal portfolio weights, $\phi_{AA}(\epsilon)$ is the shrinkage factor[16] and w_{MIN} and w_{MV} are the portfolio weights of the mean-variance portfolio and the minimum-variance portfolio respectively. In this special case, GUW show that the optimal portfolio weights can be written as a weighted average of the portfolio weights of the mean-variance portfolio and the minimum-variance portfolio, as is precisely the case for the Bayes-Stein shrinkage estimator. The degree of shrinkage depends on the precision of the return estimates and on the ambiguity parameter ϵ. As ϵ increases, the optimal portfolio is shifted stronger towards the minimum-variance portfolio. GUW, however, show that the Bayes-Stein approach is only a special case of the ambiguity-averse model if ambiguity is estimated jointly for all assets and not in case ambiguity about expected returns is estimated separately for different subsets of assets.

[12] To account for estimation errors regarding the variance-covariance matrix, Frost and Savarino (1986) and Ledoit and Wolf (2003) use a shrinkage estimator not only for expected returns, but also for the variance-covariance matrix.

[13] See Garlappi et al. (2007, p. 48).

[14] See Garlappi et al. (2007) Proposition 1 for the optimization problem if ambiguity about expected returns is estimated asset by asset, Proposition 2 if ambiguity about expected returns is estimated for all assets jointly and Proposition 3 if ambiguity about expected returns is estimated for subsets of assets.

[15] See Garlappi et al. (2007, p. 61).

[16] For the derivation of the shrinkage factor, please see Appendix A.4.

If the investor wants to consider both, parameter and model uncertainty, the ambiguity about expected returns needs to be estimated separately for different subsets of assets. In this case, estimates about expected returns are determined by a certain factor model (e.g. the CAPM) and the investor is uncertain about the assumed return-generating model to be the true model. With regard to the model incorporating parameter and model uncertainty, GUW consider one ambiguity parameter for the risky assets and a separate one for the factors. The parameter ϵ_a represents the ambiguity about the expected asset returns, while the parameter ϵ_b represents the ambiguity about the expected returns of the factor portfolios. GUW show that if the ambiguity about the expected asset returns is relatively high, compared to the ambiguity about the expected returns of the factor portfolios, the investor prefers a portfolio that puts a higher weight on the factor portfolios compared to the risky assets and vice versa.

As argued earlier, GUW show, that if parameter uncertainty only is considered, their model shows the same structure as the Bayes-Stein approach. Therefore, Scherer (2007) argues that robust optimization models that incorporate ambiguity aversion, don't add any value to Bayesian shrinkage estimation because the set of optimal portfolios doesn't change. He states that the models don't differ from each other since the optimal portfolio weights for both models are a weighted average of the portfolio weights of the mean-variance portfolio and the minimum-variance portfolio. The only difference lies in the shrinkage factor, which again depends on the ambiguity parameter ϵ. Scherer criticizes, that GUW don't state how to determine this parameter or how it is related to the risk aversion parameter γ. Consequently, the decision maker needs to form a subjective belief about the ambiguity parameter ϵ. Although GUW fix the degree of ambiguity aversion to 1, the investor still needs to determine the degree of ambiguity subjectively. Furthermore, assuming a higher degree of ambiguity about expected returns has the same effect as increasing the degree of ambiguity aversion in the model of GUW, since the ambiguity parameter ϵ is defined as the product of ambiguity and ambiguity aversion.

Since the portfolio weights depend strongly on the choice of the ambiguity parameter ϵ, it is important to find an objective measure for ambiguity and to separate ambiguity and attitudes toward it[17]. One example for an objective measure of ambiguity was introduced by Izhakian (2012). He suggests to measure ambiguity by the variance of probabilities with a parameter ranging from 0, given that all probabilities are perfectly known to 1 in case the probability is either 0 or 1 with equal probabilities. This measure of ambiguity makes it also possible to rank alternatives according to their degree of ambiguity.

[17]See Klibanoff et al. (2005), for a model which separates between ambiguity and ambiguity aversion. As mentioned in Chapter 3.4.4, it can be shown that separating preferences from beliefs leads to a better fit of the data, however, the parameters in this model also need to be determined subjectively.

4.2. Empirical results

4.2.1. Empirical setting

GUW test their ambiguity-averse approach against various other models in the literature, like the classical mean-variance approach, the minimum-variance approach and the Bayes-Stein approach. As a performance measure, they use the Sharpe ratio and therefore they test whether even a decision maker with mean-variance preferences can achieve a higher utility by taking ambiguity into account. To measure the utility of an ambiguity averse investor, it is necessary to characterize ambiguity averse preferences by a utility function, in which the utility increases as ambiguity decreases. Izhakian (2012) developed an expected utility framework that incorporates ambiguity, in which expected utility increases as the expected return increases, and utility decreases as the standard deviation increases or the degree of ambiguity increases. Izhakian further develops an extended measure of the Sharpe ratio, that incorporates ambiguity. This is a better measure for an ambiguity averse investor to compare different portfolio optimization strategies, since investors are willing to accept a lower Sharpe ratio if the portfolio ambiguity decreases (Koziol et al., 2011).

GUW use returns on eight international equity indices from January 1970 to July 2001 to measure the out-of-sample performance of the different portfolio optimization strategies. The portfolio weights are determined using historical returns from a rolling window of 120 months and the risk aversion parameter γ is set to 1. A risk aversion parameter of 1 corresponds to a low degree of risk aversion, since empirical measures have shown that the degree of risk aversion typically lies in a range between 2 and 4[18]. This low degree of risk aversion contradicts the very conservative approach of GUW and is not consistent with the extreme shrinkage towards the minimum-variance portfolio (Scherer, 2007). Scherer (2007) argues, that this contradiction results from the separation of risk preferences and ambiguity preferences, since the optimal portfolio depends strongly on the relation between risk aversion and ambiguity aversion. Therefore, it doesn't seem realistic to assume a low degree of risk aversion and at the same time a high value for the ambiguity parameter ϵ. To show the effect of the choice of the risk aversion parameter γ, the empirical study of GUW is replicated using a different value for γ in Chapter 5.3.2.

The empirical results of GUW are divided into two parts. One part considering the model accounting for parameter uncertainty only and the second part, in which the model that additionally takes model uncertainty into account is considered.

4.2.2. Parameter uncertainty only

With regard to GUW's first empirical application, it is accounted for parameter uncertainty only and the ambiguity about all assets is estimated jointly. Therefore, the optimal portfolio weights can be expressed as (13). The investor can

[18]See for example Paravisini et al. (2016)

build his portfolio from eight international equity indices and a risk-free asset doesn't exist. When comparing the Sharpe ratios of the mean-variance strategy, the minimum-variance strategy, the Bayes-Stein strategy and the ambiguity-averse approach, the authors show that the portfolios incorporating ambiguity aversion have a higher mean, lower variance and therefore higher Sharpe ratio compared to the classical mean-variance portfolio. For a sufficient high degree of ambiguity ($\epsilon > 0.25$), the ambiguity-averse approach also outperforms the Bayes-Stein model. This result, however, follows from the good performance of the minimum-variance portfolio for this specific data set, which shows the highest mean, lowest variance and therefore also the highest Sharpe ratio of all portfolios. Since the ambiguity-averse portfolio is a weighted average of the mean-variance portfolio and the minimum-variance portfolio, and the portfolio is strongly shifted towards the minimum-variance portfolio for a sufficient high degree of the ambiguity parameter ϵ, this explains the good performance of the ambiguity-averse approach.

In the empirical study of GUW, the minimum-variance portfolio shows the best out-of-sample performance, this, however, is not true in general. The data set GUW used for their empirical study has specific characteristic, e.g. returns are very noisy. Because returns are noisy, using this data set leads to an optimal strategy, that does not take expected returns into account at all. Therefore, these data-specific characteristics put an advantage to the minimum-variance approach. In Chapter 5.4, it is shown that for a different data set, it is not optimal to only hold the minimum-variance portfolio. This also reduces the superiority of the ambiguity-averse approach, compared to the classical mean-variance approach. Because of specific characteristics of historical data, Scherer (2007) argues that applying a rolling window analysis to a historical sample path differs from real out-of-sample testing. Therefore, it is important to vary certain assumptions to test the robustness of results based on historical data. A variation of important model assumptions is performed in Chapter 5.3 and in Chapter 5.4 the portfolio optimization problem is applied to a different data set.

The authors also show that the portfolios incorporating ambiguity aversion are more stable over time and show less extreme portfolio weights compared to the mean-variance portfolio and the Bayes-Stein portfolio. The higher the ambiguity parameter ϵ, the more stable are the portfolio weights over time. This is because a high parameter ϵ implies a large confidence interval resulting in a more conservative approach. In Chapter 5, a measure for portfolio turnover is introduced to directly compare the portfolio stability of different portfolio optimization strategies.

4.2.3. Parameter and model uncertainty

Regarding GUW's second empirical application, it is accounted for both, parameter and model uncertainty, and ambiguity is estimated separately for different subsets of assets. In this case, GUW assume that returns are generated by a factor model and investors are uncertain about the validity of the return-generating model. In their empirical study, GUW assume the return-generating model to be the CAPM and therefore the factor portfolio is the world market portfolio. In this model, the investor has the possibility to invest in eight international equity indices and in the factor portfolio, but again the investor is not allowed to invest in the risk-free asset. Similar to the model of Pástor (2000), discussed in Chapter 2.4, the investor's belief about the validity of the CAPM is reflected by the parameter ω, where $\omega = 0$ means the investor does not believe in the factor model and estimates expected asset returns from the historical sample and $\omega = 1$ means the investor believes dogmatically in the model and he estimates expected asset returns from the factor model. The expected return of the factor portfolio is estimated from the historical sample in both cases. In this application, the investor considers ambiguity about the estimated asset returns ϵ_a and about the estimated returns of the factor portfolio ϵ_b. If ϵ_a and ϵ_b are both 0, the resulting optimal portfolio is the mean-variance portfolio in case of $\omega = 0$ and the market portfolio in case of $\omega = 1$. GUW show that a small degree of ambiguity about the asset returns and the returns of the factor portfolio leads to an improved out-of-sample performance compared to the mean-variance portfolio and the Bayes-Stein portfolio. For large values of ϵ_a and ϵ_b, however, the performance of the ambiguity-averse portfolio declines.

Although the model incorporates model uncertainty, the authors consider only the extreme cases where the investor doesn't believe in the model at all ($\omega = 0$) or where the investor believes dogmatically in the model ($\omega = 1$). Jacquier and Polson (2010) argue that most investors have beliefs somewhere in between. Avramov (2003) states that it is neither optimal to base a portfolio on dogmatic beliefs in a return-generating model nor to completely ignore the insights from the model. He argues that, combining both sources of information, the historical sample and the return-generating model, results in higher out-of-sample Sharpe ratios. Furthermore, in GUW's approach, the investor has to make a subjective decision about his belief in the return-generating model. Instead of subjective beliefs, the parameter ω could be determined empirically, for example using posterior odd ratios or Bayesian model averaging if more than one return-generating model is considered, as discussed in Chapter 2.4.

Chapter 4 introduced the model of portfolio selection incorporating ambiguity aversion developed by GUW. In Chapter 5, the out-of-sample performance of different portfolio optimization strategies is compared with a special focus on investors who are ambiguity averse.

5. Empirical applications of the ambiguity-averse approach

5.1. Empirical design

The empirical analysis consists of three different parts. In the first part, the empirical design of GUW is adopted and their results are replicated with focus on differences in results. Both models are analyzed, the one incorporating parameter uncertainty only and the one that takes additionally

model uncertainty into account. In the first portfolio optimization problem, with parameter uncertainty only, the investor can build his portfolio from eight international equity indices (MSCI Italy, MSCI Canada, MSCI France, MSCI Germany, MSCI Japan, MSCI Switzerland, MSCI UK and MSCI USA) and a risk-free asset doesn't exist. With regard to the second portfolio optimization problem, incorporating parameter and model uncertainty, the investor additionally has the possibility to invest in the factor portfolio (MSCI World). In this second case, in order to determine the expected excess returns, the month-end US 30-day T-bill is subtracted from the expected returns. Therefore, a risk-free asset exists but it is not possible to invest in the risk-free asset. The data is computed from Datastream and from the Center for Research in Security Prices (CRSP) for the timespan from January 1970 to July 2001. The timespan is identical to the one proposed by GUW. Following the approach of GUW, a rolling window of 120 months is used to determine the optimal portfolio weights for the 121st month. These portfolio weights are then used to compute the portfolio return in the 121st month. The resulting out-of-sample period includes the timespan from January 1980 to July 2001 (259 observations). In Chapter 5.3.1, the window length is varied to test the impact of the length of the sample size on the portfolio performance and in Chapter 5.3.3 the time period is extended to December 2016.

In all empirical analyses, the models of GUW are tested against the classical mean-variance model, the minimum-variance model, the Bayes-Stein model and against the model of naïve diversification, referred to as the 1/N rule. These are the same models used by GUW, except for the 1/N rule. This model is included because of evidence that the 1/N rule can't be consistently outperformed by complex models like the one proposed by GUW (e.g. DeMiguel et al. (2009)). To compare the performance of the different portfolio optimization strategies, the Sharpe ratio and the turnover ratio are computed. The Sharpe ratio is derived as follows:

$$\text{Sharpe ratio} = \frac{\mu - r_f}{\sigma}. \tag{14}$$

The risk-free rate r_f exists, however, only with regard to the second model incorporating parameter and model uncertainty for the CAPM to be valid. Therefore, the Sharpe ratio reduces to the mean-to-standard-deviation ratio with regard to the first model taking only parameter uncertainty into account. This approach, chosen by GUW, in the second case is rather inconsistent, since the investor is not allowed to invest in the risk-free asset, but the risk-free asset is still used to evaluate the performance and for the CAPM to be valid. This approach also makes a comparison of the portfolios including and excluding the market portfolio impossible, since different performance measures are used. Nevertheless, this approach is adopted to make sure that the replicated results are comparable with the original results of GUW. Furthermore, a significance test of Sharpe ratios, as suggested by Jobson and Korkie (1981) and corrected by Memmel (2003), is performed.

However, as already mentioned in Chapter 4.2.1, if an investor is ambiguity averse and his utility depends on the degree of ambiguity, a performance measure should be implemented that takes ambiguity into account. Since the Sharpe ratio only considers risk and return, it doesn't seem plausible to expect ex ante a higher Sharpe ratio for portfolios that account for ambiguity aversion. If, nevertheless, the Sharpe ratio is used as a performance measure, it is tested whether the investor can achieve a higher out-of-sample performance by taking ambiguity into account, even if he has mean-variance preferences.

In addition to the Sharpe ratio, the portfolio turnover is computed to measure the stability of the portfolio weights. In practice, investors prefer stable portfolios that don't have to be reallocated heavily every month because they take transaction costs into account. The turnover ratio is calculated as follows[19]

$$\text{Turnover} = \frac{1}{T-n}\sum_{t=1}^{T-n}\sum_{j=1}^{N}(|\hat{w}_{j,t+1} - \hat{w}_{j,t^+}|) \tag{15}$$

in which $T-n$ is the length of the out-of-sample period, with T being the total length of the data series and n being the length of the sample size, N is the number of assets, $\hat{w}_{j,t^+}$ are the portfolio weights just before rebalancing and $\hat{w}_{j,t+1}$ are the portfolio weights right after rebalancing at time $t+1$. The turnover ratio measures the shift in portfolio weights every period and is therefore also referred to as relative turnover.

After replicating the results of GUW in Chapter 5.2, different parameters like the sample size, the degree of risk aversion and the timespan are varied in Chapter 5.3. By varying the input parameters, it is measured whether the out-of-sample performance of the different portfolio optimization strategies is robust to changes in the empirical design. Furthermore, the robustness of the superiority of the ambiguity-averse approach compared to other portfolio optimization models, as found by GUW, is tested. Finally, in Chapter 5.4, the investment universe is changed and the portfolio optimization problem is applied to the German DAX30 companies.

5.2. Replication of the empirical results of Garlappi, Uppal and Wang

5.2.1. Parameter uncertainty only

The first part of the empirical analysis deals with the replication of the empirical results of GUW. For the first empirical application, the ambiguity-averse model incorporating parameter uncertainty only is tested against the classical mean-variance model, the minimum-variance model, the Bayes-Stein model and the model of naïve diversification. With regard to this portfolio optimization problem, the investor

[19] See DeMiguel et al. (2009, p. 1929).

has the possibility to invest in eight international equity indices and a risk-free asset doesn't exist. The derivation of the portfolio weights for each strategy can be found in Appendix A. The optimal portfolio weights for the portfolios incorporating ambiguity aversion are determined from (13). Using (13), different values of the ambiguity parameter ϵ ranging from 0 to infinity are applied. However, GUW don't give any guidance on how ϵ can be determined. Therefore, the choice of ϵ remains a subjective choice of the investor. Just as done by GUW, two cases are considered, one, where short selling is allowed (Panel A) and one, where short selling is not allowed (Panel B). The original results and the replication of the results are presented in Table 1.

The replicated results are very similar to the original results of GUW. The main difference is that the mean of the mean-variance strategy in Panel A is somewhat higher, while the mean of the minimum-variance strategy in both Panels is slightly below the one found by GUW. This reduces the superiority of the minimum-variance strategy over the mean-variance strategy, but the Sharpe ratio of the minimum-variance strategy is still considerably higher and therefore, the findings presented by GUW are still observable.

The ambiguity-averse portfolios show in both Panels a higher Sharpe ratio compared to the mean-variance portfolio and compared to the Bayes-Stein portfolio, given a degree of ambiguity of 0.5 or higher. If ϵ is about 0.25, the shrinkage factor of the ambiguity-averse portfolio is approximately equivalent to the Bayes-Stein shrinkage factor while for higher values of ϵ, the shrinkage factor of the ambiguity-averse portfolio increases. In Figure 1, the shrinkage factor of the Bayes-Stein portfolio $\phi_B S$ and the shrinkage factors of two portfolios incorporating ambiguity aversion $\phi_A A$ with $\epsilon = 1$ and $\epsilon = 3$, in case that short selling is allowed, are compared[20]. From Figure 1 it can be seen that the portfolios incorporating ambiguity aversion are shrunk much more towards the minimum-variance portfolio and the shrinkage factors are more stable over time. This result is consistent with the results found by GUW. The strong shrinkage towards the minimum-variance portfolio explains the superior performance to the Bayes-Stein approach for high values of ϵ.

From Table 1, however, it can be seen that the out-of-sample performance of the ambiguity-averse approach depends strongly on the degree of ambiguity ϵ. It is therefore necessary to measure ambiguity objectively and to separate between ambiguity (belief) and ambiguity aversion (taste). Only ambiguity aversion should depend on the subjective preference of the decision maker. Without any guidance on how to determine ϵ, it can't be shown that a portfolio optimization strategy incorporating ambiguity aversion outperforms other models, like the mean-variance model or the Bayes-Stein model, in general.

A model of naïve diversification was additionally tested. The out-of-sample Sharpe ratio of the uniformly diversified portfolio, in case short selling is allowed, is higher than the Sharpe ratio of the mean-variance portfolio and the Bayes-Stein portfolio and is only outperformed by the minimum-variance portfolio and by the portfolios incorporating ambiguity aversion for extreme high levels of ambiguity ($\epsilon > 3$). The 1/N rule, however, is outperformed by the Bayes-Stein approach and the ambiguity-averse approach in case short selling is not allowed and a sufficient high amount of ambiguity is considered ($\epsilon \geq 0.25$). If short selling is prohibited, the difference in out-of-sample Sharpe ratios decreases, since imposing a short sale constraint improves the out-of-sample performance of the mean-variance portfolio, as discussed in Chapter 2.2. At the same time, the performance of the minimum-variance portfolio changes only little since the portfolio weights are less extreme and the Sharpe ratio of the uniformly diversified portfolio doesn't change at all. Consequently, the superiority of the minimum-variance approach and the 1/N rule decreases if short selling is prohibited.

Table 2 shows the results of the Jobson-Korkie significance test of Sharpe ratios for the different portfolio optimization strategies when short selling is allowed. While the Sharpe ratios of the ambiguity-averse portfolios (with $\epsilon = 1$ and $\epsilon = 3$) are significantly higher than the Sharpe ratios of the mean-variance portfolio and the Bayes-Stein portfolio on the 5%-confidence level, they don't differ significantly from the Sharpe ratio of the minimum-variance portfolio. This is because the ambiguity-averse portfolios are shrunk strongly towards the minimum-variance portfolio and therefore the portfolio structures are very similar. The Sharpe ratio of the mean-variance strategy is not significantly different from the Sharpe ratio of the minimum-variance strategy on the 5%-confidence level, however, this is true for the 10%-confidence level.[21]

From the significance test, it can be observed that the Sharpe ratios of the ambiguity-averse portfolio with $\epsilon = 3$ and the uniformly diversified portfolio are very similar, since the value of the z-statistic is very low (p-value $= 0.9656$). This indicates that the performance of the simple strategy of naïve diversification does not differ significantly from the performance of a complex model incorporating ambiguity, even if a high degree of ambiguity is considered. Note, however, that the degree of ambiguity is not taken into account in the performance measure. Therefore, this conclusion is only true for an investor with mean-variance preferences. In order to test the validity of the results for an investor who is ambiguity averse, it is necessary to develop a performance measure which accounts for ambiguity.

Additionally, to the Sharpe ratio, the turnover ratio was computed as a performance measure. In both cases, whether short selling is allowed or not, the portfolio turnover is highest for the mean-variance portfolio and lowest for the uniformly diversified portfolio. Besides the uniformly diversified

[20] The derivation of the Bayes-Stein shrinkage factor can be found in Appendix A.3, while the derivation of the shrinkage factor of the ambiguity-averse portfolios can be found in Appendix A.4.

[21] However, Jobson and Korkie (1981) state that the statistical power of the Jobson-Korkie significance test is low. Thus, a z-statistic which indicates that the Sharpe ratios are significantly different is a strong evidence.

Table 1: Replication with parameter uncertainty only

This table reports the out-of-sample mean-to-standard-deviation ratio and turnover ratio for the returns on different portfolio optimization strategies. The parameter ϵ presents the ambiguity parameter and varies between 0 and infinity. The values for the means and the standard deviations are expressed as percentage per month. Turnover is expressed as shift in portfolio weights per month. The degree of risk aversion γ is assumed to be 1.

Strategy	Original results of GUW			Replication of results			
	Mean	SD	Mean/SD	Mean	SD	Mean/SD	Turnover
Panel A: Short sales allowed							
Mean-Variance	0.0049	0.2557	0.0190	0.0064	0.2550	0.0252	2.0584
Minimum-Variance	0.0118	0.0419	0.2827	0.0109	0.0421	0.2589	0.0723
Bayes-Stein	0.0071	0.1058	0.0671	0.0066	0.1203	0.0552	0.9298
Naive Diversification	-	-	-	0.0097	0.0439	0.2214	0.0000
Ambiguity-averse							
$\epsilon = 0.00$	0.0049	0.2557	0.0190	0.0064	0.2550	0.0252	2.0584
$\epsilon = 0.25$	0.0065	0.1307	0.0495	0.0068	0.1323	0.0515	1.0343
$\epsilon = 0.50$	0.0080	0.0953	0.0841	0.0080	0.0981	0.0813	0.7136
$\epsilon = 0.75$	0.0091	0.0765	0.1190	0.0088	0.0796	0.1112	0.5411
$\epsilon = 1.00$	0.0098	0.0654	0.1491	0.0094	0.0682	0.1375	0.4374
$\epsilon = 1.50$	0.0104	0.0545	0.1909	0.0099	0.0562	0.1755	0.3208
$\epsilon = 2.00$	0.0107	0.0501	0.2144	0.0101	0.0511	0.1977	0.2582
$\epsilon = 2.50$	0.0109	0.0480	0.2281	0.0103	0.0487	0.2107	0.2211
$\epsilon = 3.00$	0.0111	0.0467	0.2369	0.0104	0.0473	0.2189	0.1973
$\epsilon \to \infty$	0.0118	0.0419	0.2827	0.0109	0.0421	0.2589	0.0723
Panel B: Short sales not allowed							
Mean-Variance	0.0104	0.0587	0.1774	0.0103	0.0587	0.1752	0.2072
Minimum-Variance	0.0117	0.0412	0.2831	0.0107	0.0415	0.2568	0.0481
Bayes-Stein	0.0106	0.0511	0.2074	0.0101	0.0435	0.2322	0.1090
Naive Diversification	-	-	-	0.0097	0.0439	0.2214	0.0000
Ambiguity-averse							
$\epsilon = 0.00$	0.0104	0.0587	0.1774	0.0103	0.0587	0.1752	0.2072
$\epsilon = 0.25$	0.0113	0.0511	0.2214	0.0101	0.0446	0.2270	0.1183
$\epsilon = 0.50$	0.0115	0.0482	0.2391	0.0103	0.0427	0.2402	0.0906
$\epsilon = 0.75$	0.0115	0.0467	0.2472	0.0104	0.0421	0.2458	0.0774
$\epsilon = 1.00$	0.0116	0.0457	0.2533	0.0104	0.0419	0.2487	0.0702
$\epsilon = 1.50$	0.0116	0.0446	0.2607	0.0105	0.0417	0.2516	0.0629
$\epsilon = 2.00$	0.0117	0.0440	0.2647	0.0105	0.0416	0.2531	0.0592
$\epsilon = 2.50$	0.0117	0.0436	0.2671	0.0106	0.0416	0.2539	0.0570
$\epsilon = 3.00$	0.0117	0.0433	0.2695	0.0106	0.0416	0.2544	0.0556
$\epsilon \to \infty$	0.0117	0.0412	0.2831	0.0107	0.0415	0.2568	0.0481

portfolio, which does not show any relative turnover at all[22], the minimum-variance portfolio has the lowest turnover ratio and the turnover ratios of the Bayes-Stein portfolio and the ambiguity-averse portfolios lie between the turnover ratio of the mean-variance portfolio and the minimum-variance portfolio. Since the shrinkage factors of the ambiguity-averse portfolios for a sufficient high level of ϵ are higher than the shrinkage factor of the Bayes-Stein portfolio, this explains the lower turnover of portfolios accounting for ambiguity aversion compared to the Bayes-Stein portfolio. If short selling is not allowed, the turnover of the mean-variance portfolio declines strongly, but is still considerably higher than the turnover of the minimum-variance portfolio. This reduction of turnover can again be explained by the improvement of the mean-variance strategy by imposing a short selling constraint. If short sales are prohibited, portfolio weights are less extreme and the portfolio turnover decreases. A graphical representation of the fluctuation of portfolio weights for the different portfolio optimization strategies, as was also presented by GUW, can be found in Appendix B.

[22]The uniformly diversified portfolio, however, shows absolute turnover. The portfolio weights don't change over time, but absolute prices change and therefore also the 1/N rule requires monthly reallocation (DeMiguel et al., 2009).

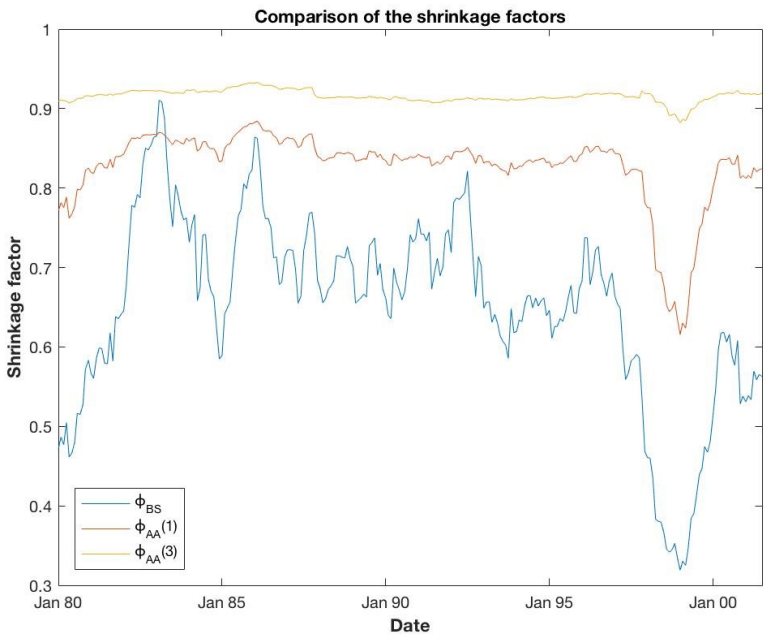


Figure 1: Comparison of the shrinkage factors
This figure compares the shrinkage factors of the Bayes-Stein portfolio and two portfolios incorporating ambiguity aversion. The blue line shows the shrinkage factor of the Bayes-Stein approach, the red line shows the shrinkage factor if the ambiguity parameter ϵ is set to 1 and the yellow line shows the shrinkage factor if the ambiguity parameter ϵ is set to 3. The degree of risk aversion γ is assumed to be 1.

Table 2: Jobson-Korkie significance test – replication

This table reports the values of the z-statistics according to the significance test of Sharpe ratios suggested by Jobson and Korkie (1981) and corrected by Memmel (2003). The Sharpe ratios of the mean-variance strategy (MV), the minimum-variance strategy (MIN), the Bayes-Stein strategy (BS), the 1/N rule (ND), the ambiguity-averse approach with $\epsilon = 1$ ($\epsilon = 1.00$) and the ambiguity-averse approach with $\epsilon = 3$ ($\epsilon = 3.00$) are tested. One star * means that the p-value indicates that the Sharpe ratios are significantly different on the 5%-confidence level, while two stars ** indicate that the Sharpe ratios are significantly different on the 1%-confidence level.

Jobson-Korkie significance test

Strategy	MV	MIN	BS	ND	$\epsilon = 1.00$	$\epsilon = 3.00$
Mean-Variance	1.0000	-1.9474	-0.9388	-1.6444	-2.2130*	-2.3380*
Minimum-Variance		1.0000	1.8858	-0.8964	1.4216	0.8728
Bayes-Stein			1.0000	1.5252	-2.8397**	-2.3352*
Naive Diversification				1.0000	0.9438	0.0432
$\epsilon = 1.00$					1.0000	-1.8133
$\epsilon = 3.00$						1.0000

5.2.2. Parameter and model uncertainty

With regard to the second empirical application, the ambiguity-averse approach incorporating both, parameter and model uncertainty, is tested against the mean-variance model, the minimum-variance model, the Bayes-Stein model and the model of naïve diversification. GUW also include Pastor's data and model approach presented in Chapter 2.4, but since it is nested in the ambiguity-averse approach (in case ϵ_a and ϵ_b are both 0), this approach isn't explicitly considered here. With regard to this portfolio optimization problem, the investor has the possibility to invest in the market portfolio in addition to the eight international equity indices. For the CAPM to be valid, a risk-free asset is introduced, but the investor is not allowed to invest in the risk-free asset.

The optimal portfolio weights for the ambiguity-averse portfolios are determined from the equation system (A.5), which can be found in Appendix A.5. Using (A.5), different values for the ambiguity about the expected asset returns ϵ_a and about the expected returns of the factor portfolio ϵ_b are applied. The equation system (A.5) doesn't restrict the portfolio weights to sum to 1 and, at the same time, the investor is not allowed to invest in the risk-free asset. Therefore, in order to solve for the optimal portfolio weights, the resulting weights for the risky assets w_a and the factor portfolio w_b need to be upscaled to one. This procedure seems to be implausible because, if the portfolio weights are very small or

very large, they are strongly modified to make sure that the investor invests his whole wealth in the stock market. Furthermore, if the ambiguity about the expected asset returns ϵ_a is very high and at the same time the ambiguity about the expected returns of the market portfolio ϵ_b is very high, the investor might not want to participate in the stock market at all. This result of limited market participation, in case that ambiguity is sufficiently high, was discussed in Chapter 3.5.1. GUW, however, don't give any guidance on how the investor should behave if the degree of ambiguity is so high that he prefers not to participate in the stock market at all.

The original results of the portfolio optimization problem with parameter and model uncertainty and the replicated results are shown in Table 3. Following the approach of GUW, the tables are divided into two parts. One, where the investor does not believe in the CAPM ($\omega = 0$) and the expected excess returns of the assets $\hat{\mu}_a$ are estimated from the historical sample, and one, where the investor believes dogmatically in the CAPM ($\omega = 1$) and $\hat{mu}_a$ is estimated from the CAPM in which $\hat{mu}_a = \bar{\beta}\hat{\mu}_b$ and $\bar{\beta}$ is the N x 1 vector of betas. The expected excess return of the market portfolio $\hat{\mu}_b$ is estimated from the historical sample in both cases.

From Table 3 it can be seen, that GUW show in their results that the out-of-sample Sharpe ratio of the mean-variance portfolio and the Bayes-Stein portfolio can be improved by allowing for a moderate level of ambiguity. However, the Sharpe ratio of the ambiguity-averse portfolio declines if ϵ_a and ϵ_b are large. If ϵ_a is relatively high compared to ϵ_b, the investor prefers a portfolio which puts high weights on the world market portfolio. In the extreme case, the optimal portfolio equals the world market portfolio and the Sharpe ratio is 0.1239. If, however, ϵ_b is relatively high compared to ϵ_a, the portfolio is shifted towards a portfolio consisting only of the risky assets. This effect is stronger in case the investor does not believe in the CAPM ($\omega = 0$), while the portfolio structure, in case the investor believes dogmatically in the CAPM ($\omega = 1$) depends only little on ϵ_a, whereas it is strongly affected by changes in ϵ_b.

The results of the replication show a different, however, intuitive picture of the out-of-sample performance of the different portfolio optimization strategies. The first difference to be noted is that the Sharpe ratios of the mean-variance portfolio and the Bayes-Stein portfolio are considerably higher compared to the original results of GUW. For the replication, when applying these two strategies, it was assumed that the investor has the possibility to invest in nine different assets, the eight international equity indices plus the world market portfolio and a separation between the risky assets and the world market portfolio didn't take place. The missing separation when applying the mean-variance strategy, also explains why the Sharpe ratio of the mean-variance portfolio differs from the Sharpe ratio of the ambiguity-averse portfolio with $\epsilon_a = 0$ and $\epsilon_b = 0$.

The main difference between the original results and the replicated results, however, is the difference in Sharpe ratios of the portfolios incorporating ambiguity aversion. If $\epsilon_b = 0$ and $\epsilon_a > 0$, the investor is uncertain about the returns of the risky assets, while he is not uncertain about the returns of the market portfolio. This leads to a shift of the optimal portfolio towards the world market portfolio. The replicated results show that in case $\omega =$

0, already an increase in ϵ_a from 0 to 0.25 leads to an optimal portfolio, which equals the market portfolio and the Sharpe ratio is 0.1156. Given that $\epsilon_b = 0$ and $\omega = 1$, the investor prefers to invest all his wealth in the world market portfolio, even if he is certain about the asset returns ($\epsilon_a = 0$). The shift towards the world market portfolio in case that $\epsilon_b = 0$ and $\epsilon_a > 0$ is consistent with the results of GUW. However, in case that $\omega = 0$, the original results show that the optimal portfolio equals the world market portfolio only for sufficient high values of ϵ_a (≥ 2.5) and otherwise the optimal portfolio consists of both, the risky assets and the world market portfolio.

If, on the other hand, $\epsilon_a = 0$ and $\epsilon_b > 0$, the investor is certain about the returns of the risky assets, but uncertain about the returns of the world market portfolio. This leads to a shift of the optimal portfolio towards a portfolio consisting only of the risky assets. Again, an increase in the ambiguity parameter ϵ_b from 0 to 0.5 is sufficient in order for the investor to prefer not to invest in the world market portfolio at all and instead to invest all his wealth in the risky assets. This is true for both cases, $\omega = 0$ and $\omega = 1$. However, in case the investor believes dogmatically in the CAPM, the Sharpe ratios are considerably higher. This result arises because the portfolio weights are less extreme if expected asset returns are determined from the CAPM instead of estimating them from the historical sample. Because of estimation errors, these extreme portfolio weights lead to a poor out-of-sample performance. The shift towards a portfolio consisting of only risky assets is consistent with the original results of GUW, the Sharpe ratios, however, differ strongly.

In the general case of $\epsilon_a > 0$ and $\epsilon_b > 0$, the investor is uncertain about the asset returns and also uncertain about the returns of the world market portfolio. The replicated results show that this leads to a situation, in which the investor doesn't want to participate in the stock market at all and instead he prefers to invest all his wealth in the risk-free asset. This result is confirmed by Boyle et al. (2012), who apply the ambiguity-averse approach as proposed by GUW and show that, if ambiguity reaches a certain level, the investor prefers to invest only in the risk-free asset. This preference, however, contradicts the construction of the portfolio optimization problem incorporating parameter and model uncertainty presented by GUW, because they don't allow for an investment in the risk-free asset. This is especially inconsistent, because a risk-free asset is introduced to determine the excess returns and for the CAPM to be valid, but the risk-free asset is not an investment option. As a result, the investor doesn't invest any wealth and the Sharpe ratio is set to 0. This result differs from the original result of GUW, who show that in the general case of $\epsilon_a > 0$ and $\epsilon_b > 0$, the investor chooses a portfolio that consists of both, the risky assets and the world market portfolio.

GUW showed in their empirical study that, most of the

Table 3: Replication with parameter and model uncertainty

This table reports the out-of-sample Sharpe ratio for the returns on different portfolio optimization strategies. The parameter ϵ_a presents the ambiguity about the asset returns and varies between 0 and 3 and ϵ_b presents the ambiguity about the returns of the factor portfolio, also varying between 0 and 3. The parameter ω presents the investor's beliefs in the validity of the CAPM, where $\omega = 0$ means the investor doesn't believe in the CAPM and $\omega = 1$ means the investor believes dogmatically in the CAPM. The values for the Sharpe ratios are expressed as percentage per month.

Results of Garlappi, Uppal and Wang							
Strategy	Sharpe ratio						
Mean-Variance	-0.0719						
Minimum-Variance	0.1490						
Bayes-Stein	-0.0528						
Naive Diversification	-						
Ambiguity-averse	With $\omega = 0$			ϵ_b			
	0.00	0.50	1.00	1.50	2.00	2.50	3.00
$\epsilon_a = 0.00$	-0.0853	0.1127	0.1127	0.1127	0.1127	0.1127	0.1127
$\epsilon_a = 0.25$	-0.0774	0.1032	0.1043	0.1043	0.1043	0.1043	0.1043
$\epsilon_a = 0.50$	-0.0475	0.0824	0.0815	0.0824	0.0839	0.0842	0.0842
$\epsilon_a = 0.75$	-0.0113	0.0604	0.0568	0.0524	0.0477	0.0489	0.0481
$\epsilon_a = 1.00$	0.0930	0.0655	0.0601	0.0566	0.0562	0.0556	0.0447
$\epsilon_a = 1.50$	0.1219	0.1218	0.0814	0.0503	0.0391	0.0319	-0.0435
$\epsilon_a = 2.00$	0.1223	0.1252	0.0874	0.0643	0.0487	0.0341	-0.0492
$\epsilon_a = 2.50$	0.1239	0.1275	0.0909	0.0465	0.0126	0.0091	-0.0870
$\epsilon_a = 3.00$	0.1239	0.1284	0.0943	0.0551	0.0197	0.0364	-0.0647
	With $\omega = 1$						
$\epsilon_a = 0.00$	0.1239	0.1202	0.1202	0.1202	0.1202	0.1202	0.1202
$\epsilon_a = 0.25$	0.1239	0.1284	0.0943	0.0551	0.0197	0.0383	0.0383
$\epsilon_a = 0.50$	0.1239	0.1284	0.0943	0.0551	0.0197	0.0364	-0.0647
$\epsilon_a = 0.75$	0.1239	0.1284	0.0943	0.0551	0.0197	0.0364	-0.0647
$\epsilon_a = 1.00$	0.1239	0.1284	0.0943	0.0551	0.0197	0.0364	-0.0647
$\epsilon_a = 1.50$	0.1239	0.1284	0.0943	0.0551	0.0197	0.0364	-0.0647
$\epsilon_a = 2.00$	0.1239	0.1284	0.0943	0.0551	0.0197	0.0364	-0.0647
$\epsilon_a = 2.50$	0.1239	0.1284	0.0943	0.0551	0.0197	0.0364	-0.0647
$\epsilon_a = 3.00$	0.1239	0.1284	0.0943	0.0551	0.0197	0.0364	-0.0647
Replication of results							
Strategy	Sharpe ratio						
Mean-Variance	0.0238						
Minimum-Variance	0.1295						
Bayes-Stein	0.0261						
Naive Diversification	0.0994						
Ambguity-averse	With $\omega = 0$			ϵ_b			
	0.00	0.50	1.00	1.50	2.00	2.50	3.00
$\epsilon_a = 0.00$	-0.0616	0.0295	-0.0295	-0.0295	-0.0295	-0.0295	-0.0295
$\epsilon_a = 0.25$	0.1156	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\epsilon_a = 0.50$	0.1156	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\epsilon_a = 0.75$	0.1156	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\epsilon_a = 1.00$	0.1156	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\epsilon_a = 1.50$	0.1156	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\epsilon_a = 2.00$	0.1156	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\epsilon_a = 2.50$	0.1156	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\epsilon_a = 3.00$	0.1156	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

(Continued)

Table 3—continued

	With $\omega = 1$						
$\epsilon_a = 0.00$	0.1156	0.0950	0.0950	0.0950	0.0950	0.0950	0.0950
$\epsilon_a = 0.25$	0.1156	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\epsilon_a = 0.50$	0.1156	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\epsilon_a = 0.75$	0.1156	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\epsilon_a = 1.00$	0.1156	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\epsilon_a = 1.50$	0.1156	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\epsilon_a = 2.00$	0.1156	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\epsilon_a = 2.50$	0.1156	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\epsilon_a = 3.00$	0.1156	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

time, allowing for ambiguity in portfolio optimization improves the out-of-sample performance, even if ambiguity is not included in the performance measure. In the following Chapter, several input parameters like the sample size, the degree of risk aversion and the timespan are varied to test the robustness of the results. In Chapter 5.3, only the model with parameter uncertainty is considered, because the model accounting additionally for model uncertainty always leads to an extreme result, where the investor either invests all his wealth in the world market portfolio, he invests all his wealth in the risky assets or he doesn't participate in the stock market at all.

5.3. Variation of the input parameters

5.3.1. Sample size

The first input parameter that is varied is the sample size, which is used to determine the optimal portfolio weights. GUW use a window length of 120 months. In this Chapter, the window length is varied between 30 months and 120 months. A maximum of 120 months is used, to make sure that the out-of-sample period remains unchanged. The case is considered, where there is parameter uncertainty only and short selling is allowed. Figure 2 presents the resulting Sharpe ratios (Panel A) and the turnover ratios (Panel B) for the different values of the window length.

Panel A shows that, for every window length, the Sharpe ratio of the minimum-variance portfolio is the highest, while the Sharpe ratio of the mean-variance portfolio is the lowest. Furthermore, it can be observed, that the ranking of the portfolio optimization strategies regarding the Sharpe ratio for every window length is the same. Also, there is no clear trend visible for either strategy as the window length increases. These results indicate that the original results of GUW are robust to changes in the sample size.

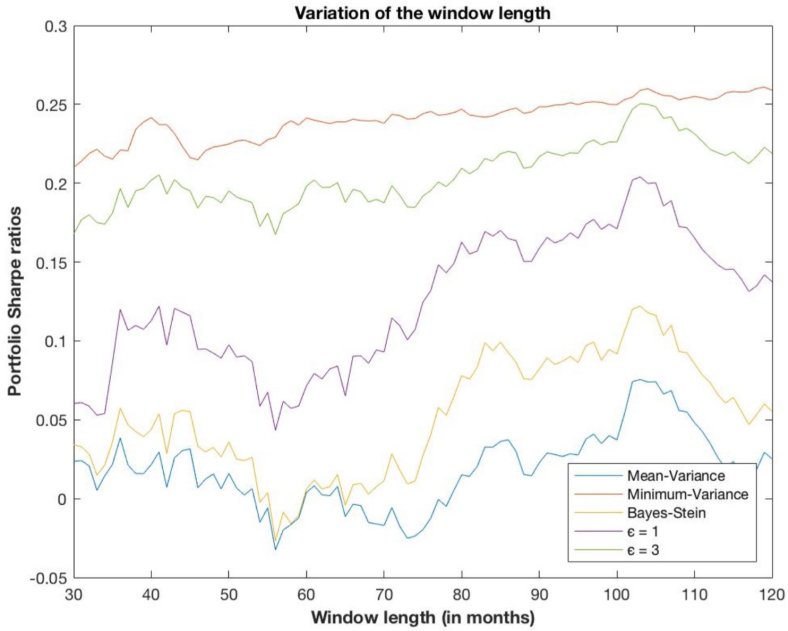
From Panel B, a clear trend becomes visible that for every strategy the portfolio turnover decreases as the window length increases. However, the effect is greatest for the mean-variance portfolio. If the sample size is large, many historical observations are used to determine the expected return and therefore, portfolio weights are less extreme and the turnover ratio decreases. The minimum-variance portfolio has the lowest turnover ratio which decreases only to a small extent as the sample size increases. This is because the minimum-variance portfolio is independent of expected returns, but also the variance-co-

variance matrix is more stable if many observations are taken into account. Consequently, the turnover ratio is not robust to the choice of the window length as it can be reduced if the window length is increased. However, the ranking of the different portfolio optimization strategies regarding the turnover ratio does not change and is consistent with the original assumption of 120 months.

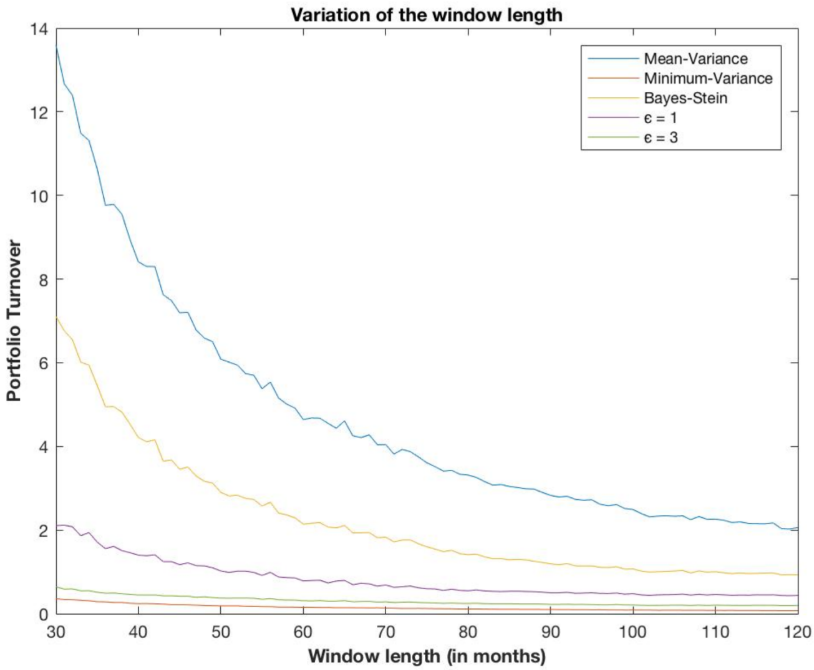
5.3.2. The degree of risk aversion

The investor's degree of risk aversion is presented by the variable γ. While GUW assume that γ is fixed at 1, empirical evidence has shown that most investors have a degree of risk aversion between 2 and 4. Therefore, a value of 1 reflects a low degree of risk aversion. It has been shown that incorporating ambiguity aversion to the portfolio optimization problem leads to more conservative portfolios that are less risky than the mean-variance portfolio. This effect, however, can also be achieved by increasing the degree of risk aversion. If γ increases, the investor prefers a portfolio further to the left on the efficient frontier, closer to the minimum-variance portfolio. Furthermore, the degree of ambiguity aversion is correlated with the degree of risk aversion (Uppal and Wang, 2003, Bossaerts et al., 2010). Scherer (2007) even argues, that it is impossible to distinguish between risk aversion and ambiguity aversion. GUW fix the degree of ambiguity aversion also at 1, however, they don't give any guidance on how the ambiguity parameter ϵ can be determined. Therefore, ϵ depends on the investor's subjective level of ambiguity about expected returns. Consequently, the assumption that investors are strongly uncertain about expected returns (high values of ϵ) contradicts their low degree of risk aversion.

Since most investors have a degree of risk aversion between 2 and 4, the empirical study of GUW is repeated assuming that $\gamma = 3$. This assumption corresponds to an investor who has a moderate degree of risk aversion and therefore prefers a more conservative portfolio. In this empirical application, the case with parameter uncertainty only is considered. Table 4 presents the out-of-sample performance of the mean-variance strategy, the minimum-variance strategy,



Panel A: Sharpe ratios



Panel B: Portfolio Turnover

Figure 2: Variation of the window length
This figure presents the Sharpe ratios (Panel A) and the turnover ratios (Panel B) for different portfolio optimization strategies using window lengths between 30 months and 120 months. The portfolio strategies considered are the mean-variance strategy (blue line), the minimum-variance strategy (red line), the Bayes-Stein strategy (yellow line), the ambiguity-averse approach with $\epsilon = 1$ (purple line) and the ambiguity-averse approach with $\epsilon = 3$ (green line).

the Bayes-Stein strategy, the ambiguity-averse approach and the strategy of naïve diversification. Panel A shows the results if short selling is allowed, Panel B the results if short selling is not allowed.

An increase in γ does not affect the out-of-sample performance of the minimum-variance portfolio and the uniformly diversified portfolio, since they are independent of γ. However, in Panel A, an increase in γ leads to an increased performance of the mean-variance portfolio, as the mean increases while the standard deviation decreases strongly, leading to an improved out-of-sample Sharpe ratio. Additionally, the turnover ratio of the mean-variance portfolio is reduced sharply, because portfolio weights are less extreme as γ increases. These effects also lead to an increase in Sharpe ratio and decrease in turnover ratio of the Bayes-Stein portfolio. As γ gets larger, the effect of increasing the ambiguity parameter reduces, since the difference in performance between the mean-variance port-

folio and the minimum-variance portfolio gets smaller. As a result, increasing γ, when applying the classical mean-variance strategy, has a very similar effect to increasing ϵ when applying the ambiguity-averse approach. Both adjustments lead to an optimal portfolio which is shifted from the mean-variance portfolio with $\gamma = 1$ towards the minimum-variance portfolio. If γ approaches infinity, the optimal portfolio equals the minimum-variance portfolio, as is precisely the case if ϵ approaches infinity.

If short sales are not allowed, the out-of-sample performance of the mean-variance portfolio increases only slightly as γ increases. This is because imposing a short sale constraint has a strong effect on the performance of the mean-variance portfolio as discussed in Chapter 2.2 and Chapter 5.2.1. Since the additional constraint leads to better diversified portfolios with less extreme portfolio weights, an increase in γ, which has a similar effect, improves the out-of-sample performance only little.

Table 5 presents the Jobson-Korkie significance test of Sharpe ratios, which indicates whether the Sharpe ratios of the different portfolio optimization strategies change significantly if γ is increased from 1 to 3. The table considers only the case when short selling is allowed, since the Sharpe ratio reacts only little to an increase in γ, if a short sale constraint is imposed.

From Table 5 it can be seen that the Sharpe ratio of the mean-variance portfolio increases significantly on the 1%-confidence level if γ is increased from 1 to 3, while the Sharpe ratio of the Bayes-Stein portfolio increases significantly on the 5%-confidence level. The Sharpe ratios of the ambiguity-averse portfolios with $\epsilon = 1$ and $\epsilon = 3$, however, don't increase significantly. This result confirms, that increasing γ has a similar effect on the out-of-sample performance to taking ambiguity aversion into account, when optimizing portfolios. The significance test, however, also reveals that the portfolios considering ambiguity aversion are more robust to changes in the degree of risk aversion.

Finally, Figure 3 compares the shrinkage factors of the Bayes-Stein portfolio and the ambiguity-averse portfolios with $\epsilon = 1$ and $\epsilon = 3$ in case short selling is allowed.

If γ increases from 1 to 3, the shrinkage factor of the Bayes-Stein portfolio and the shrinkage factors of the portfolios incorporating ambiguity aversion are increasingly similar. This explains why the superiority of the ambiguity-averse portfolios over the Bayes-Stein portfolio, decreases as γ increases. However, from Figure 3 it can be seen that the shrinkage factors of the ambiguity-averse portfolios are much more stable and that the fluctuation of the shrinkage factor decreases as ϵ increases. This result is in accordance with the results presented in Figure 1 for a degree of risk aversion of 1.

If the degree of risk aversion increases, the out-of-sample performance of the mean-variance strategy and the Bayes-Stein strategy increases significantly, while the performance of the ambiguity-averse approach increases only slightly. As a result, the superiority of the ambiguity-averse approach decreases as γ increases. On the other hand, it was shown that the ambiguity-averse approach is much more robust to changes in the risk aversion parameter.

5.3.3. Time period

GUW use in their empirical study observations from January 1970 to July 2001. In this Chapter, the timespan is extended to December 2016, while the investment universe and all other original input parameters remain unchanged. The extension of the time period indicates whether the ambiguity-averse approach outperforms other strategies, like the mean-variance approach and the Bayes-Stein approach, also if a different, more current time period is considered. As already argued in Chapter 4.2.2, historical sample paths have specific characteristics that might put an advantage to a certain model. If this advantage arises from time-specific characteristics, varying the time period is an appropriate method to test the robustness of the results over time. Firstly, a single extended time period from January 1970 to December 2016 with 564 observations is analyzed and afterwards a rolling out-of-sample period of 259 observations between January 1970 and December 2016 is considered.

Table 6 presents the out-of-sample performance of the mean-variance strategy, the minimum-variance strategy, the Bayes-Stein strategy, the ambiguity-averse approach and the strategy of naïve diversification for the time period from January 1970 to December 2016 when parameter uncertainty only is considered. Only the case when short sales are allowed is analyzed, because, analogous to the original time period, the performance of the different strategies is very similar if short sales are prohibited. Table 7 presents the Jobson-Korkie significance test of Sharpe ratios for the extended time period.

From Table 6, it can be seen that the minimum-variance portfolio outperforms all other portfolios regarding mean, standard deviation and Sharpe ratio also if the time period is extended to December 2016. However, the superiority of the minimum-variance portfolio decreases, especially regarding the mean. Table 7 shows that the Sharpe ratio of the minimum-variance portfolio isn't significantly higher than the Sharpe ratio of the mean-variance portfolio on the 5%-confidence level. The z-statistic increased from -1.9474 to -1.5978 compared to the original time period, confirming that the superiority of the minimum-variance portfo-

Table 4: Variation of the degree of risk aversion

This table reports the out-of-sample mean-to-standard-deviation ratio and turnover ratio for the returns on different portfolio optimization strategies. The parameter ϵ presents the ambiguity parameter and varies between 0 and infinity. The values for the means and the standard deviations are expressed as percentage per month. Turnover is expressed as shift in portfolio weights per month. The degree of risk aversion γ is assumed to be 3.

Variation of the degree of risk aversion

Strategy	Mean	SD	Mean/SD	Turnover
Panel A: Short sales allowed				
Mean-Variance	0.0094	0.0935	0.1006	0.6913
Minimum-Variance	0.0109	0.0421	0.2589	0.0723
Bayes-Stein	0.0095	0.0558	0.1700	0.3189
Naive Diversification	0.0097	0.0439	0.2214	0.0000
Ambiguity-averse				
$\epsilon = 0.00$	0.0094	0.0935	0.1006	0.6913
$\epsilon = 0.25$	0.0098	0.0644	0.1516	0.4110
$\epsilon = 0.50$	0.0100	0.0574	0.1738	0.3333
$\epsilon = 0.75$	0.0101	0.0537	0.1885	0.2877
$\epsilon = 1.00$	0.0102	0.0513	0.1990	0.2568
$\epsilon = 1.50$	0.0103	0.0486	0.2126	0.2173
$\epsilon = 2.00$	0.0104	0.0472	0.2209	0.1927
$\epsilon = 2.50$	0.0105	0.0462	0.2265	0.1757
$\epsilon = 3.00$	0.0105	0.0456	0.2305	0.1633
$\epsilon \to \infty$	0.0109	0.0421	0.2589	0.0723
Panel B: Short sales not allowed				
Mean-Variance	0.0103	0.0510	0.2014	0.1668
Minimum-Variance	0.0107	0.0415	0.2568	0.0481
Bayes-Stein	0.0103	0.0425	0.2427	0.0860
Naive Diversification	0.0097	0.0439	0.2214	0.0000
Ambiguity-averse				
$\epsilon = 0.00$	0.0103	0.0510	0.2014	0.1668
$\epsilon = 0.25$	0.0103	0.0444	0.2326	0.1074
$\epsilon = 0.50$	0.0104	0.0433	0.2394	0.0923
$\epsilon = 0.75$	0.0104	0.0428	0.2430	0.0838
$\epsilon = 1.00$	0.0104	0.0426	0.2452	0.0783
$\epsilon = 1.50$	0.0105	0.0422	0.2478	0.0715
$\epsilon = 2.00$	0.0105	0.0421	0.2493	0.0674
$\epsilon = 2.50$	0.0105	0.0420	0.2504	0.0646
$\epsilon = 3.00$	0.0105	0.0419	0.2511	0.0625
$\epsilon \to \infty$	0.0107	0.0415	0.2568	0.0481

lio decreased. The ranking of the portfolio optimization strategies regarding their out-of-sample Sharpe ratio doesn't change if the extended time period is considered and, therefore, the ambiguity-averse portfolios still outperform the mean-variance portfolio and also the Bayes-Stein portfolio for a sufficient high degree of ambiguity ($\epsilon > 0.25$). However, compared to the original time period, for which the ambiguity-averse portfolios showed a significantly higher Sharpe ratio than the mean-variance portfolio on the 5%-confidence level, this isn't true anymore for the extended time period. In general, the results from Chapter 5.2.1 can be confirmed if the time period is extended, however, the effect of accounting for ambiguity aversion is stronger for the original time period.

Additionally, to extending the time period, a rolling out-of-sample period of 259 observations is considered. Beginning with the original starting date in January 1970, the starting date is shifted by one month until the end of the time period in December 2016 is reached. This approach results in 186 different out-of-sample periods with 259 observations each. The Sharpe ratios of the different portfolio optimization strategies for the rolling out-of-sample periods are presented in Figure 4.

Figure 4 shows, that the Sharpe ratio of the minimum-variance portfolio exceeds the Sharpe ratio of the mean-variance portfolio for every out-of-sample period considered.

Table 5: Jobson-Korkie significance test for different degrees of risk aversion

This table reports the values of the z-statistics according to the significance test of Sharpe ratios suggested by Jobson and Korkie (1981) and corrected by Memmel (2003). The significance of Sharpe ratios for the returns on different portfolio optimization strategies for different degrees of risk aversion ($\gamma = 1$ and $\gamma = 3$) is measured. The Sharpe ratios of the mean-variance strategy (MV), the Bayes-Stein strategy (BS), the ambiguity-averse approach with $\epsilon = 1$ ($\epsilon = 1.00$) and the ambiguity-averse approach with $\epsilon = 3$ ($\epsilon = 3.00$) are tested. One star * means that the p-value indicates that the Sharpe ratios are significantly different on the 5%-confidence level, while two stars ** indicate that the Sharpe ratios are significantly different on the 1%-confidence level.

Jobson-Korkie significance test

			$\gamma = 3$		
	Strategy	MV	BS	$\epsilon = 1.00$	$\epsilon = 3.00$
	Mean-Variance	-2.7077**			
$\gamma = 1$	Bayes-Stein		-2.5082*		
	$\epsilon = 1.00$			-1.8255	
	$\epsilon = 3.00$				-1.3722

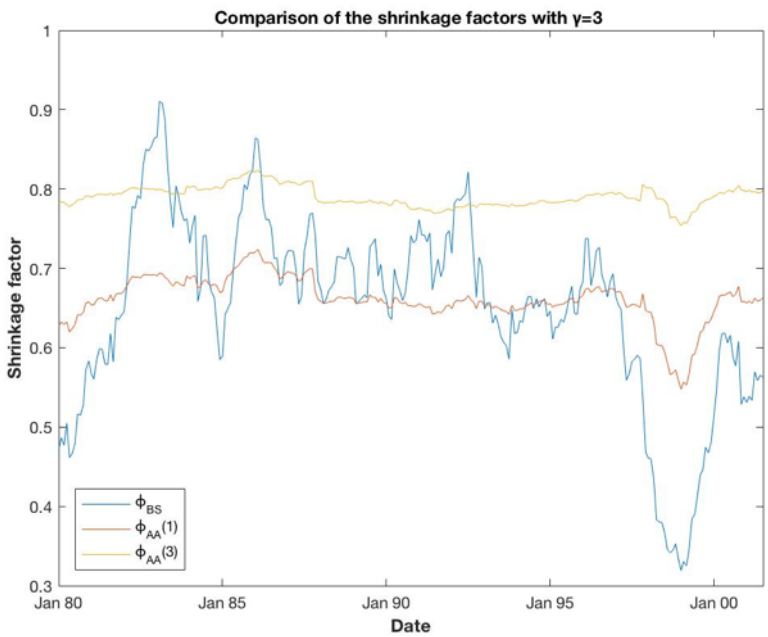


Figure 3: Comparison of the shrinkage factors with $\gamma = 3$
This figure compares the shrinkage factors of the Bayes-Stein portfolio and two portfolios incorporating ambiguity aversion. The blue line shows the shrinkage factor of the Bayes-Stein approach, the red line shows the shrinkage factor if the ambiguity parameter ϵ is set to 1 and the yellow line shows the shrinkage factor if the ambiguity parameter ϵ is set to 3. The degree of risk aversion γ is assumed to be 3.

This confirms the result, that estimation errors regarding expected returns are so high, that the performance can be improved by ignoring them. In general, the results from Figure 4 are in line with the results from the previous Chapters. It is, however, important to notice that although the minimum-variance portfolio shows the highest Sharpe ratio in most cases, this is not true for every out-of-sample period. The minimum-variance approach is not able to consistently outperform the 1/N rule, while the ambiguity-averse portfolio only shows a higher Sharpe ratio than the minimum-variance portfolio for very high values of ϵ and only for few out-of-sample periods. From Figure 4 it can also be observed that the Sharpe ratios of the minimum-variance portfolio, the uniformly diversified portfolio and the ambiguity-averse portfolio with $\epsilon = 3$ for most out-of-sample periods are very similar. It is intuitive that the Sharpe ratios of the minimum-variance portfolio and the ambiguity-averse portfolio with $\epsilon = 3$ are similar, since the ambiguity-averse portfolio is strongly shrunk towards the minimum-variance portfolio if ϵ is large. The strong performance of the uniformly diversified portfolio, however, indicates that estimation errors regarding expected asset moments are so large, that an approach that completely ignores the historical sample can't be consistently outperformed by more sophisticated models.

In Chapter 5.3, different input parameters were varied to test the robustness of the original results presented by GUW. It was shown that the ambiguity-averse approach outperforms the mean-variance strategy and the Bayes-Stein strat-

Table 6: Variation of the time period

This table reports the out-of-sample mean-to-standard-deviation ratio and turnover ratio for the returns on different portfolio optimization strategies. The parameter ϵ presents the ambiguity parameter and varies between 0 and infinity. The values for the means and the standard deviations are expressed as percentage per month. Turnover is expressed as shift in portfolio weights per month. The degree of risk aversion γ is assumed to be 1. The timespan considered extends from January 1970 to December 2016.

Variation of the time period

Strategy	Mean	SD	Mean/SD	Turnover
Mean-Variance	0.0074	0.2729	0.0272	2.6701
Minimum-Variance	0.0086	0.0401	0.2150	0.0878
Bayes-Stein	0.0060	0.1353	0.0444	1.3517
Naive Diversification	0.0074	0.0468	0.1589	0.0000
Ambiguity-averse				
$\epsilon = 0.00$	0.0074	0.2729	0.0272	2.6701
$\epsilon = 0.25$	0.0063	0.1477	0.0426	1.4684
$\epsilon = 0.50$	0.0067	0.1079	0.0621	1.0450
$\epsilon = 0.75$	0.0072	0.0849	0.0850	0.7798
$\epsilon = 1.00$	0.0076	0.0709	0.1073	0.6052
$\epsilon = 1.50$	0.0080	0.0570	0.1407	0.4145
$\epsilon = 2.00$	0.0082	0.0512	0.1602	0.3245
$\epsilon = 2.50$	0.0083	0.0484	0.1716	0.2748
$\epsilon = 3.00$	0.0084	0.0468	0.1790	0.2433
$\epsilon \to \infty$	0.0086	0.0401	0.2150	0.0878

Table 7: Jobson-Korkie significance test - extended time period

This table reports the values of the z-statistics according to the significance test of Sharpe ratios suggested by Jobson and Korkie (1981) and corrected by Memmel (2003). The Sharpe ratios of the mean-variance strategy (MV), the minimum-variance strategy (MIN), the Bayes-Stein strategy (BS), the 1/N rule (ND), the ambiguity-averse approach with $\epsilon = 1$ ($\epsilon = 1.00$) and the ambiguity-averse approach with $\epsilon = 3$ ($\epsilon = 3.00$) are tested. One star * means that the p-value indicates that the Sharpe ratios are significantly different on the 5%-confidence level, while two stars ** indicate that the Sharpe ratios are significantly different on the 1%-confidence level. The timespan considered extends from January 1970 to December 2016.

Jobson-Korkie significance test

Strategy	MV	MIN	BS	ND	$\epsilon = 1.00$	$\epsilon = 3.00$
Mean-Variance	1.0000	-1.5978	-0.6807	-1.0927	-1.7562	-1.9342
Minimum-Variance		1.0000	1.5949	-1.0656	1.2720	0.7659
Bayes-Stein			1.0000	1.0185	-2.2208*	-2.0054*
Naive Diversification				1.0000	0.5471	-0.2936
$\epsilon = 1.00$					1.0000	-1.7001
$\epsilon = 3.00$						1.0000

egy (for $\epsilon > 0.5$) regarding all settings. The superiority of the ambiguity-averse approach, however, decreases as the degree of risk aversion increases and if the time period is extended to December 2016. The ambiguity-averse approach was only able to outperform the minimum-variance approach for very few settings, when a different out-of-sample period was considered. Moreover, the ambiguity-averse portfolio only showed a higher Sharpe ratio than the uniformly diversified portfolio for very high values of ϵ. In accordance with the original results, the minimum-variance portfolio showed the highest mean, lowest standard deviation and consequently also the highest Sharpe ratio for almost all settings. Since the portfolios, incorporating a sufficient high degree of ambiguity aversion, are strongly shrunk towards the minimum-variance portfolio, this explains their good performance in all settings. Furthermore, it was shown that the performance of the ambiguity-averse approach is more robust to changes in the empirical design than the mean-variance approach and the Bayes-Stein approach. The performance of the minimum-variance strategy and the 1/N rule, however, is even more robust to changes in input parameters, especially because they're both independent of the degree of risk aversion. The 1/N rule is additionally independent of the sample size.

While in this Chapter several input parameters were varied, the investment universe remained unchanged. The following Chapter deals with a portfolio optimization problem, in which the investor has the possibility to invest in the German DAX30 stocks. By changing the investment universe,

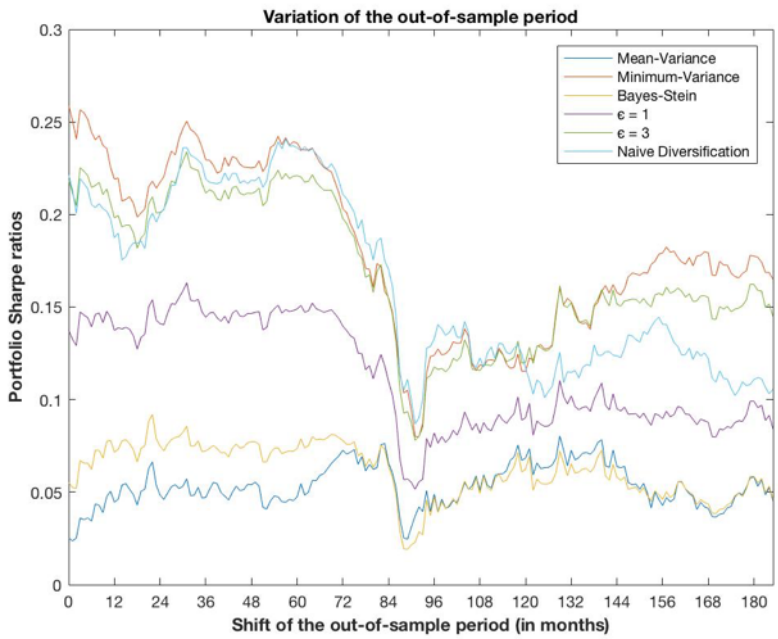


Figure 4: Variation of the out-of-sample period
This figure presents the Sharpe ratios for the returns on different portfolio optimization strategies for 186 different out-of-sample periods. The starting date varies between January 1970 and July 1985. The x-axis shows how many months the starting date differs from the original starting date in January 1970. The sample size to determine the portfolio weights is assumed to be 120 months. The portfolio strategies considered are the mean-variance strategy (dark blue line), the minimum-variance strategy (red line), the Bayes-Stein strategy (yellow line), the ambiguity-averse approach with $\epsilon = 1$ (purple line), the ambiguity-averse approach with $\epsilon = 3$ (green line) and the 1/N rule (light blue line

the robustness of the different portfolio optimization strategies is tested when the special characteristics of the original investment universe don't apply.

5.4. Application to the DAX30 stocks

5.4.1. Empirical setting

In this part of the empirical analysis, the investment universe is changed and the investor builds his portfolio from the German DAX30 stocks. The returns for the time period January 1973 to December 2016 are considered. From the 30 DAX30 companies, only 16 companies were a constituent of the DAX30 over the whole time period. Consequently, the investment universe consists of these 16 German stocks (Allianz, BASF, BMW, Bayer, Beiersdorf, Commerzbank, Continental, Deutsche Bank, E-ON, Heidelberg Cement, Linde, Deutsche Lufthansa, Münchener Rück, RWE, Siemens, ThyssenKrupp). All other assumptions are taken from GUW. Portfolio weights are determined using a rolling window of 120 months and the degree of risk aversion is set to 1. Using the original assumptions of GUW, a comparison between the original results for an investment universe of eight international equity indices and the new investment universe of 16 DAX30 stocks is straightforward.

If the investment universe is shifted from eight international equity indices to the German DAX30 stocks, two main characteristics of the portfolio optimization problem change. Firstly, instead of equity indices, now single stocks are considered. The returns of single stocks fluctuate much stronger compared to the returns of equity indices, because single stocks have a higher idiosyncratic risk in addition to the systematic market risk. As a result, diversification has a stronger effect if single stocks are considered. The second main difference is the focus on a single country. This focus leads to country-specific risks that can't be diversified away. Consequently, the standard deviation of a portfolio consisting only of stocks from one country is higher compared to an internationally diversified portfolio. On the other hand, the focus on a single country might reduce ambiguity, as argued in Chapter 3.5.1. If German investors are considered, they might be more certain about expected returns of German companies compared to international companies. Therefore, if investors are sufficiently ambiguity averse, they prefer to invest in the German market only. As a result, the focus of the investor on his home country is in line with the ambiguity aversion literature (Epstein and Miao, 2003). Koziol et al. (2011) show in their empirical study, that ambiguity aversion plays an even more important role for German investors compared to international investors. All these effects are important to be considered when the portfolio optimization problem is applied to the German market.

5.4.2. Empirical results

This part presents the empirical results regarding the portfolio optimization problem in which the investor has the possibility to invest in 16 DAX30 stocks. Table 8 reports the out-of-sample performance of the mean-variance strategy,

the minimum-variance strategy, the Bayes-Stein strategy, the strategy introduced by GUW taking ambiguity aversion into account and the strategy of naïve diversification. The focus of the analysis lies on the model with parameter uncertainty only and again only the case when short selling is allowed is considered, because differences in performance are small if short selling is prohibited.

Table 8 shows that the standard deviation of each strategy is higher compared to the original dataset. This result is in accordance with the theoretical intuition that portfolios that contain stocks from only one country are riskier compared to an internationally diversified portfolio and that portfolios put together from single stocks are riskier than portfolios of equity indices. It can also be observed that the portfolio turnover of all strategies is higher compared to the original dataset. This is because the returns fluctuate more heavily if single stocks are considered and therefore, the portfolio needs to be reallocated more often.

From Table 8 it can be seen that, for this dataset, the mean-variance portfolio shows the highest mean, while the minimum-variance portfolio shows the lowest standard deviation. In contrast to the original dataset, for which the minimum-variance portfolio had the highest mean, the investor now can gain higher returns by taking additional risk. This is exactly the relationship Markowitz depicted with the efficient frontier. However, to gain higher returns, the investor needs to bear a much higher risk compared to the minimum-variance portfolio and as a result, the Sharpe ratio of the minimum-variance portfolio is higher than the Sharpe ratio of the mean-variance portfolio. The difference in Sharpe ratios, however, is not significant on the 5%-confidence level as can be seen from the Jobson-Korkie significance test presented in Table 9. The Bayes-Stein portfolio has a mean which is almost as high as the one of the mean-variance portfolio, but the standard deviation is lower, leading to a higher out-of-sample Sharpe ratio. The Bayes-Stein portfolio is equivalent to a portfolio incorporating ambiguity aversion with an ambiguity parameter of about 0.25. As a result, taking high values of ambiguity into account ($\epsilon > 0.25$) improves the out-of-sample performance due to the strong shift towards the minimum-variance portfolio, which shows the highest Sharpe ratio. As in the original dataset proposed by GUW, the portfolio turnover of the minimum-variance portfolio is much lower compared to the mean-variance portfolio, because the mean-variance portfolio exhibits extreme portfolio weights and requires massive reallocations.

In comparison to the original dataset, it can be seen from Table 8 that the performance of the 1/N rule compared to the ambiguity-averse approach decreased. While for the original dataset an ambiguity parameter larger than 3 was necessary to outperform the 1/N rule regarding the Sharpe ratio, the ambiguity parameter must only equal 1 in this example. This is intuitive because portfolios of individual stocks have a higher idiosyncratic volatility than portfolios of equity indices and therefore the performance of the 1/N rule decreases.

Table 9 shows that, according to the Jobson-Korkie significance test, the out-of-sample Sharpe ratio of no strategy is significantly higher on the 5%-confidence level than the Sharpe ratio of any other strategy. Although the statistical power of the Jobson-Korkie significance test is low, it becomes clear that the superiority of the minimum-variance strategy and the ambiguity-averse approach, decreased, compared to the original dataset.

The empirical application to the German DAX30 stocks was also implemented for GUW's model incorporating both, parameter and model uncertainty. In this implementation, it is assumed that assets follow a factor structure and the investor is uncertain about the return-generating model to be the true model. The return-generating model is assumed to be the CAPM and therefore the factor portfolio is the market portfolio. With regard to this portfolio optimization problem, the investor has the possibility to invest in the market portfolio in addition to the 16 DAX30 stocks. However, similar to the original dataset presented in Chapter 5.2.2, the investor prefers not to participate in the stock market at all if he is uncertain about the stock returns ($\epsilon_a > 0$) and also uncertain about the returns of the market portfolio ($\epsilon_b > 0$). If he is only uncertain about the stock returns, he invests all his wealth in the market portfolio (for $\epsilon_a > 0.25$ in case $\omega = 0$ and for $\epsilon_a \geq 0$ in case $\omega = 1$) and if he is only uncertain about the returns of the market portfolio, he invests in a portfolio consisting of only DAX30 stocks (for $\epsilon_b > 0$). Only if $\epsilon_a \leq 0.25$, $\epsilon_b = 0$ and $\omega = 0$, the investor prefers a mixed portfolio of DAX30 stocks and the market portfolio. Consequently, allowing for ambiguity about the stock returns and the returns of the market portfolio, leads to limited market participation. Since it is not allowed to invest in the risk-free asset, the Sharpe ratio in case $\epsilon_a > 0$ and $\epsilon_b > 0$ is set to 0 and the ambiguity-averse approach is outperformed by all other portfolio optimization strategies, which show a positive Sharpe ratio. The detailed results are presented in Appendix C.

The empirical application of the portfolio optimization problem to the DAX30 stocks revealed that, in general, the original results of GUW can be confirmed. Allowing for ambiguity leads to a higher out-of-sample Sharpe ratio compared to the mean-variance strategy and the Bayes-Stein strategy (for $\epsilon > 0.25$). This is, because the ambiguity-averse approach shrinks the mean-variance portfolio stronger towards the minimum-variance portfolio than the Bayes-Stein approach and the minimum-variance portfolio showed the highest Sharpe ratio. Also with regard to the DAX30 stocks, the ambiguity-averse approach wasn't able to outperform the minimum-variance portfolio, while the performance of the ambiguity-averse approach increased compared to the 1/N rule. Moreover, when looking at the model incorporating both, parameter and model uncertainty, ambiguity aversion has such a strong effect that the investor prefers not to participate in the stock market at all. To sum up, the results of the application to the DAX30 stocks are in line with the results of the original dataset, but the strong performance of the minimum-variance approach and the ambiguity-averse approach is weakened.

Table 8: Application to the DAX30 stocks

This table reports the out-of-sample mean-to-standard-deviation ratio and turnover ratio for the returns on different portfolio optimization strategies. The parameter ϵ presents the ambiguity parameter and varies between 0 and infinity. The values for the means and the standard deviations are expressed as percentage per month. Turnover is expressed as shift in portfolio weights per month. The degree of risk aversion γ is assumed to be 1. The timespan considered extends from January 1973 to December 2016.

Application to the DAX30 stocks

Strategy	Mean	SD	Mean/SD	Turnover
Mean-Variance	0.0128	0.4538	0.0282	3.9405
Minimum-Variance	0.0087	0.0550	0.1590	0.1693
Bayes-Stein	0.0127	0.2058	0.0616	1.6240
Naive Diversification	0.0073	0.0635	0.1147	0.0000
Ambiguity-averse				
$\epsilon = 0.00$	0.0128	0.4538	0.0282	3.9405
$\epsilon = 0.25$	0.0126	0.2098	0.0603	1.6670
$\epsilon = 0.50$	0.0117	0.1388	0.0842	1.0253
$\epsilon = 0.75$	0.0107	0.1047	0.1025	0.7201
$\epsilon = 1.00$	0.0101	0.0880	0.1152	0.5664
$\epsilon = 1.50$	0.0096	0.0740	0.1296	0.4258
$\epsilon = 2.00$	0.0094	0.0684	0.1370	0.3617
$\epsilon = 2.50$	0.0093	0.0654	0.1415	0.3245
$\epsilon = 3.00$	0.0092	0.0635	0.1445	0.2999
$\epsilon \to \infty$	0.0087	0.0550	0.1590	0.1693

Table 9: Jobson-Korkie significance test – DAX30

This table reports the values of the z-statistics according to the significance test of Sharpe ratios suggested by Jobson and Korkie (1981) and corrected by Memmel (2003). The Sharpe ratios of the mean-variance strategy (MV), the minimum-variance strategy (MIN), the Bayes-Stein strategy (BS), the 1/N rule (ND), the ambiguity-averse approach with $\epsilon = 1$ ($\epsilon = 1.00$) and the ambiguity-averse approach with $\epsilon = 3$ ($\epsilon = 3.00$) are tested. One star * means that the p-value indicates that the Sharpe ratios are significantly different on the 5%-confidence level, while two stars ** indicate that the Sharpe ratios are significantly different on the 1%-confidence level. The timespan considered extends from January 1973 to December 2016.

Jobson-Korkie significance test

Strategy	MV	MIN	BS	ND	$\epsilon = 1.00$	$\epsilon = 3.00$
Mean-Variance	1.0000	-1.0932	-1.4745	-0.6345	-1.6477	-1.4114
Minimum-Variance		1.0000	0.8882	-0.6553	0.5636	0.3180
Bayes-Stein			1.0000	0.4068	-1.3921	-1.1660
Naive Diversification				1.0000	-0.0044	-0.3330
$\epsilon = 1.00$					1.0000	-0.8398
$\epsilon = 3.00$						1.0000

6. Conclusion

The classical mean-variance portfolio optimization model, introduced by Harry Markowitz in 1952, assumes that expected asset moments are given and estimation errors are ignored. Since expected returns are unknown, they are inevitably estimated with error, leading to a bad out-of-sample performance of optimized portfolios. To overcome this weakness, Bayesian approaches were introduced, that are explicitly designed to reduce estimation errors. Bayesian approaches, however, assume that the investor has a unique prior about the asset moments. In this way, it is implicitly assumed that the investor is ambiguity neutral. Empirical studies (e.g. Ellsberg (1961)), however, have shown that investors are averse to ambiguity. As a consequence, new models were developed allowing for multiple priors about expected returns and thereby accounting for ambiguity aversion. This thesis is focusing on the multi-prior approach introduced by Garlappi et al. (2007) which accounts for ambiguity regarding both, expected returns (parameter uncertainty) and the return-generating model (model uncertainty). The authors show that, using their ambiguity-averse approach, the out-of-sample performance can be improved compared to the mean-variance approach and the Bayes-Stein approach.

In Chapter 5, an empirical study was performed to test whether incorporating ambiguity aversion to portfolio optimization increases the out-of-sample performance regarding

the Sharpe ratio and turnover ratio for two different datasets. Since the performance measures don't account for ambiguity, it can only be tested, whether even an investor with mean-variance preferences can benefit from taking ambiguity into account. To measure the utility of the different portfolio optimization strategies for an investor who is ambiguity averse, it is necessary to measure ambiguity objectively. Furthermore, the performance of the ambiguity-averse approach depends strongly on the ambiguity parameter ϵ, while Garlappi, Uppal and Wang don't give any guidance on how the ambiguity parameter can be determined. Therefore, a general statement about the performance of the ambiguity-averse approach can only be given if ambiguity is measured objectively and if it is incorporated in the performance measure.

The empirical applications have shown that the original results of Garlappi, Uppal and Wang could approximately be replicated if the model with parameter uncertainty only is considered. For the model taking both, parameter and model uncertainty, into account, however, the results differ, since, in contrast to the original results, even a small degree of ambiguity about the risky assets and the factor portfolio makes the investor prefer not to participate in the stock market at all. Limited market participation in case that investors are ambiguity averse is intuitive and was observed by various authors in the literature. But since the empirical setting of Garlappi, Uppal and Wang doesn't allow the investor to invest in the risk-free asset, a conceptual contradiction arises. Consequently, the focus of the empirical study lies on the model incorporating parameter uncertainty only.

The variation of the input parameters revealed that the results of the ambiguity-averse approach, especially for high values of ϵ, are fairly robust. If the degree of risk aversion increases, the results of the ambiguity-averse approach are comparatively stable, but the superiority over the mean-variance approach and the Bayes-Stein approach decreases. This is because increasing the degree of risk aversion has a similar effect to increasing the ambiguity parameter. At the same time, it is difficult to differentiate both concepts, since they are correlated with each other. Changing the time period and the investment universe revealed that the results presented by Garlappi, Uppal and Wang are robust to time-specific and asset-specific characteristics. However, the superiority of the ambiguity-averse approach decreased in both cases.

Despite the results that show that the ambiguity-averse approach outperforms the mean-variance approach and the Bayes-Stein approach for a sufficient high level of ϵ, it is important to notice that the performance of the ambiguity-averse approach depends strongly on the performance of the minimum-variance approach. Garlappi, Uppal and Wang have shown that the ambiguity-averse portfolios are a weighted average of the mean-variance portfolio and the minimum-variance portfolio. Therefore, the approach is observationally equivalent to the Bayes-Stein approach. But since the shrinkage factor of the ambiguity-averse approach for most values of ϵ is higher, the portfolios are shrunk stronger towards the minimum-variance portfolio. Consequently, above-mentioned results are achieved due to the superiority of the minimum-variance approach regarding both datasets. It is also important to consider that the 1/N rule was outperformed by the ambiguity-averse approach for high values of ϵ only. At the same time, the turnover ratio of the uniformly diversified portfolio was always the lowest, since the portfolio must only be adjusted to price changes in the stocks and not to fluctuations in the portfolio weights.

From these results, the question arises whether simple strategies like the minimum-variance strategy or the strategy of naïve diversification should be preferred to more complex approaches, like the ambiguity-averse approach. For almost all the empirical applications, the minimum-variance portfolio showed the highest out-of-sample Sharpe ratio and the lowest standard deviation, while the turnover ratio was low. The strategy of naïve diversification outperformed the mean-variance strategy in all applications and the Bayes-Stein strategy in all applications, allowing for short selling, regarding the Sharpe ratio. Furthermore, the uniformly diversified portfolio always showed the lowest turnover ratio. DeMiguel et al. (2009) compare 14 portfolio optimization models, including the classical mean-variance model, the minimum-variance model, the Bayes-Stein model, the ambiguity-averse approach introduced by Garlappi, Uppal and Wang, and the 1/N rule and find that no model is consistently better than the 1/N rule regarding Sharpe ratio and portfolio turnover. They show that the 1/N rule can only be consistently outperformed if the sample size is very large or if the number of assets is small. These assumptions, however, contradict real-life investment situations. The performance of the 1/N rule is particularly strong, if portfolios of stocks (e.g. equity indices) are considered, as done by DeMiguel, Garlappi and Wang. It was shown in Chapter 5.4.2, that the 1/N rule, however, also performs well if individual stocks are considered.

The good performance of the minimum-variance strategy and the strategy of naïve diversification indicates that estimation errors regarding expected returns are so large, that the benefits from optimal asset allocation are overshadowed (DeMiguel et al., 2009). Regarding the ambiguity-averse approach introduced by Garlappi, Uppal and Wang, the results imply that the role of ambiguity aversion in portfolio optimization is rather small compared to the importance of estimation errors. The minimum-variance strategy and the 1/N rule are both independent of expected returns and therefore also of estimation errors regarding them. This makes both strategies easy to implement and explains their good out-of-sample performance. The 1/N rule is even completely independent of realized returns, especially also of the variance-covariance matrix, and is therefore the least complex model to implement. Since realized returns are a very poor measure of future returns, the focus in future research should lie on the improvement of the estimation of asset moments (DeMiguel et al., 2009; Elton, 1999). Consequently, the development of new measures for expected returns seems to be more promising than the development of complex portfolio optimization rules that are based on realized returns.

References

Aït-Sahali, Y. and Brandt, M. W. Variable selection for portfolio choice. *The Journal of Finance*, 56(4):1297–1351, 2001.

Al-Najjar, N. I. and Weinstein, J. The ambiguity aversion literature: a critical assessment. *Economics & Philosophy*, 25(3):249–284, 2009.

Anderson, E. W. and Cheng, A.-R. Robust bayesian portfolio choices. *The Review of Financial Studies*, 29(5):1330–1375, 2016.

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.

Anscombe, F. J., Aumann, R. J., et al. A definition of subjective probability. *Annals of Mathematical Statistics*, 34(1):199–205, 1963.

Antoniou, C., Galariotis, E. C., and Read, D. Ambiguity aversion, company size and the pricing of earnings forecasts. *European Financial Management*, 20(3):633–651, 2014.

Antoniou, C., Harris, R. D., and Zhang, R. Ambiguity aversion and stock market participation: An empirical analysis. *Journal of Banking & Finance*, 58:57–70, 2015.

Avramov, D. Stock return predictability and model uncertainty. *Journal of Financial Economics*, 64(3):423–458, 2002.

Avramov, D. Stock return predictability and asset pricing models. *The Review of Financial Studies*, 17(3):699–738, 2003.

Avramov, D. and Zhou, G. Bayesian portfolio analysis. *Annual Review of Financial Economics*, 2(1):25–47, 2010.

Banz, R. W. Editorial: zero correlation between theory and practice in financial economics? Financial Markets and Portfolio Management, 11(4), 383-391, 1997.

Barry, C. B. Portfolio analysis under uncertain means, variances, and covariances. *The Journal of Finance*, 29(2):515–522, 1974.

Best, M. J. and Grauer, R. R. On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results. *The Review of Financial Studies*, 4(2):315–342, 1991.

Binmore, K., Stewart, L., and Voorhoeve, A. How much ambiguity aversion? *Journal of Risk and Uncertainty*, 45(3):215–238, 2012.

Black, F. and Litterman, R. Global portfolio optimization. *Financial Analysts Journal*, 48(5):28–43, 1992.

Bossaerts, P., Ghirardato, P., Guarnaschelli, S., and Zame, W. R. Ambiguity in asset markets: Theory and experiment. *The Review of Financial Studies*, 23(4):1325–1359, 2010.

Boyle, P., Garlappi, L., Uppal, R., and Wang, T. Keynes meets markowitz: The trade-off between familiarity and diversification. *Management Science*, 58 (2):253–272, 2012.

Brandt, M. W. Portfolio choice problems. In *Handbook of financial econometrics: Tools and techniques: 269-336*. Elsevier, 2009.

Branger, N., Larsen, L. S., and Munk, C. Robust portfolio choice with ambiguity and learning about return predictability. *Journal of Banking & Finance*, 37(5):1397–1411, 2013.

Broadie, M. Computing efficient frontiers using estimated parameters. *Annals of Operations Research*, 45(1):21–58, 1993.

Brown, S. J. Optimal portfolio choice under uncertainty: A bayesian approach. *Ph. D. Dissertation, Graduate School of Business, University of Chichago*, 1976.

Campanale, C. Learning, ambiguity and life-cycle portfolio allocation. *Review of Economic Dynamics*, 14(2):339–367, 2011.

Cao, H. H., Wang, T., and Zhang, H. H. Model uncertainty, limited market participation, and asset prices. *The Review of Financial Studies*, 18(4): 1219–1251, 2005.

Chan, L. K., Karceski, J., and Lakonishok, J. On portfolio optimization: Forecasting covariances and choosing the risk model. *The Review of Financial Studies*, 12(5):937–974, 1999.

Chen, H., Ju, N., and Miao, J. Dynamic asset allocation with ambiguous return predictability. *Review of Economic Dynamics*, 17(4):799–823, 2014.

Chen, Z. and Epstein, L. Ambiguity, risk, and asset returns in continuous time. *Econometrica*, 70(4):1403–1443, 2002.

Chopra, V. K. and Ziemba, W. T. The effect of errors in means, variances, and covariances on optimal portfolio choice. *Journal of Portfolio Management*, 19(2):6–11, 1993.

Chopra, V. K., Hensel, C. R., and Turner, A. L. Massaging mean-variance inputs: returns from alternative global investment strategies in the 1980s. *Management Science*, 39(7):845–855, 1993.

Choquet, G. Theory of capacities. In *Annales de l'institut Fourier, Volume 5: 131-295*, 1954.

Clasing, H. and Rudd, A. Modern Portfolio Theory 2, Volume 2, California, Orinda, 1988.

Cremers, K. M. Stock return predictability: A bayesian model selection perspective. *The Review of Financial Studies*, 15(4):1223–1249, 2002.

DeMiguel, V. and Nogales, F. J. Portfolio selection with robust estimation. *Operations Research*, 57(3):560–577, 2009.

DeMiguel, V., Garlappi, L., and Uppal, R. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The Review of Financial Studies*, 22(5):1915–1953, 2009.

Deng, X.-T., Li, Z.-F., and Wang, S.-Y. A minimax portfolio selection strategy with equilibrium. *European Journal of Operational Research*, 166(1):278–292, 2005.

Dow, J. and da Costa Werlang, S. R. Uncertainty aversion, risk aversion, and the optimal choice of portfolio. *Econometrica: Journal of the Econometric Society*, pages 197–204, 1992.

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

Elton, E. J. Presidential address: expected return, realized return, and asset pricing tests. *The Journal of Finance*, 54(4):1199–1220, 1999.

Epstein, L. G. and Wang, T. Intertemporal asset pricing under knightian uncertainty. *Econometrica: Journal of the Econometric Society*, pages 283–322, 1994.

Epstein, L. G. and Miao, J. A two-person dynamic equilibrium under ambiguity. *Journal of Economic Dynamics and Control*, 27(7):1253–1288, 2003.

Epstein, L. G. and Schneider, M. Learning under ambiguity. *The Review of Economic Studies*, 74(4):1275–1303, 2007.

Epstein, L. G. and Schneider, M. Ambiguity, information quality, and asset pricing. *The Journal of Finance*, 63(1):197–228, 2008.

Etner, J., Jeleva, M., and Tallon, J.-M. Decision theory under ambiguity. *Journal of Economic Surveys*, 26(2):234–270, 2012.

Fletcher, J. and Hillier, J. An examination of resampled portfolio efficiency. *Financial Analysts Journal*, 57(5):66–74, 2001.

Frost, P. A. and Savarino, J. E. An empirical bayes approach to efficient portfolio selection. *Journal of Financial and Quantitative Analysis*, 21(3): 293–305, 1986.

Frost, P. A. and Savarino, J. E. For better performance: Constrain portfolio weights. *Journal of Portfolio Management*, 15(1):29–34, 1988.

Garlappi, L., Uppal, R., and Wang, T. Portfolio selection with parameter and model uncertainty: A multi-prior approach. *The Review of Financial Studies*, 20(1):41–81, 2007.

Gelman, A., Carlin, J., Stern, H. S., and Rubin, D. Bayesian data analysis. London: Chapman & Hall, 1995.

Gilboa, I. and Schmeidler, D. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18(2):141–153, 1989.

Gilboa, I. and Marinacci, M. Ambiguity and the bayesian paradigm. In *Readings in Formal Epistemology*, pages 385–439. Springer, 2016.

Goldfarb, D. and Iyengar, G. Robust portfolio selection problems. *Mathematics of Operations Research*, 28(1):1–38, 2003.

Grauer, R. R. and Shen, F. C. Do constraints improve portfolio performance? *Journal of Banking & Finance*, 24(8):1253–1274, 2000.

Green, R. C. and Hollifield, B. When will mean-variance efficient portfolios be well diversified? *The Journal of Finance*, 47(5):1785–1809, 1992.

Grinold, R. C. and Easton, K. K. Attribution of performance and. *Worldwide Asset and Liability Modeling*, 10:87–113, 1998.

Guidolin, M. and Rinaldi, F. Ambiguity in asset pricing and portfolio choice: A review of the literature. *Theory and Decision*, 74(2):183–217, 2013.

Hodges, S. D. and Brealey, R. A. Portfolio selection in a dynamic and uncertain world. *Financial Analysts Journal*, 28(6):50–65, 1973.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. Bayesian model averaging: a tutorial. *Statistical science*, pages 382–401, 1999.

Izhakian, Y. Capital asset pricing under ambiguity. *SSRN 2020179*, 2012.

Jacquier, E. and Polson, N. Bayesian econometrics in finance, 2010.

Jagannathan, R. and Ma, T. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, 58(4):1651–1684, 2003.

James, W. and Stein, C. Estimation with quadratic loss. Paper presented at the Proceedings of the fourth Berkeley symposium on mathematical statistics and probability, 1961.

Jobson, J. D. and Korkie, B. M. Performance hypothesis testing with the sharpe and treynor measures. *The Journal of Finance*, 36(4):889–908, 1981.

Jobson, J. D. and Korkie, B. Estimation for markowitz efficient portfolios. *Journal of the American Statistical Association*, 75(371):544–554, 1980.

Jobson, J. Improved estimation for markowitz portfolios using james-stein type estimators. In *Proceedings of the American Statistical Association, Business and Economics Statistics Section, Volume 71: 279-284*, 1979.

Jorion, P. International portfolio diversification with estimation risk. *Journal of Business*, 58(3):259–278, 1985.

Jorion, P. Bayes-stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis*, 21(3):279–292, 1986.

Kalymon, B. A. Estimation risk in the portfolio selection model. *Journal of Financial and Quantitative Analysis*, 6(1):559–582, 1971.

Kempf, A., Kreuzberg, K., and Memmel, C. How to incorporate estimation risk into markowitz optimization. In *Operations Research Proceedings 2001: 175-182*. Springer, 2002.

Klein, R. W. and Bawa, V. S. The effect of estimation risk on optimal portfolio choice. *Journal of financial economics*, 3(3):215–231, 1976.

Klibanoff, P., Marinacci, M., and Mukerji, S. A smooth model of decision making under ambiguity. *Econometrica*, 73(6):1849–1892, 2005.

Klibanoff, P., Marinacci, M., and Mukerji, S. Recursive smooth ambiguity preferences. *Journal of Economic Theory*, 144(3):930–976, 2009.

Knight, F. H. Risk, uncertainty and profit Hart, Schaffner and Marx Price Essays no.31. New York and Boston: Houghton Mifflin, 1921.

Koziol, C., Proelss, J., and Schweizer, D. Do institutional investors care about the ambiguity of their assets? evidence from portfolio holdings in alternative investments. *International Journal of Theoretical and Applied Finance*, 14(04):465–484, 2011.

Ledoit, O. and Wolf, M. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of empirical finance*, 10(5):603–621, 2003.

Lim, A. E., Shanthikumar, J. G., and Vahn, G.-Y. Robust portfolio choice with learning in the framework of regret: Single-period case. *Management Science*, 58(9):1732–1746, 2012.

Ma, X., Zhao, Q., and Qu, J. Robust portfolio optimization with a generalized expected utility model under ambiguity. *Annals of Finance*, 4(4):431–444, 2008.

Maccheroni, F., Marinacci, M., and Rustichini, A. Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica*, 74 (6):1447–1498, 2006.

Marinacci, M. Learning from ambiguous urns. *Statistical Papers*, 43(1):143–151, 2002.

Markowitz, H. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

Markowitz, H. M. Foundations of portfolio theory. *The Journal of Finance*, 46(2):469–477, 1991.

Memmel, C. Performance hypothesis testing with the sharpe ratio. *Finance Letters*, 1(1), 2003.

Merton, R. C. On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, 8(4):323–361, 1980.

Michaud, R. O. The markowitz optimization enigma: Is 'optimized'optimal? *Financial Analysts Journal*, 45(1):31–42, 1989.

Mukerji, S. and Tallon, J.-M. Ambiguity aversion and incompleteness of financial markets. *The Review of Economic Studies*, 68(4):883–904, 2001.

Nigmatullin, E. A. Bayesian model averaging for moment conditions models. Technical report, Working paper, University of Wisconsin-Madison, 2003.

Paravisini, D., Rappoport, V., and Ravina, E. Risk aversion and wealth: Evidence from person-to-person lending portfolios. *Management Science*, 63 (2):279–297, 2016.

Pástor, L. Portfolio selection and asset pricing models. *The Journal of Finance*, 55(1):179–223, 2000.

Pástor, L. and Stambaugh, R. F. Comparing asset pricing models: an investment perspective. *Journal of Financial Economics*, 56(3):335–381, 2000.

Pastor, L. and Veronesi, P. Learning in financial markets. *Annual Review of Financial Economics*, 1(1):361–381, 2009.

Pataracchia, B. Ambiguity and volatility: Asset pricing implications. 2011.

Pflug, G. and Wozabal, D. Ambiguity in portfolio selection. *Quantitative Finance*, 7(4):435–442, 2007.

Qu, X. A belief-based definition of ambiguity aversion. *Theory and Decision*, 79(1):15–30, 2015.

Raftery, A. E., Madigan, D., and Hoeting, J. A. Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, 92(437):179–191, 1997.

Raiffa, H. Risk, ambiguity, and the savage axioms: comment. *The Quarterly Journal of Economics*, 75(4):690–694, 1961.

Ramsey, F. P. Truth and probability. the foundations of mathematics and other logical essays. *Kegan, Paul, Trench, Trubner & Co., London*, pages 156–198, 1931.

Savage, L. J. The Foundations of Statistics, New York, John Wiley and Sons. 1954.

Scherer, B. *Resampling der Effizienzlinie*. Dichtl. H./Kleeberg, JM/Schlenger, C., Uhlenbruch Verlag: 319-336, 2003.

Scherer, B. Can robust portfolio optimisation help to build better portfolios? *Journal of Asset Management*, 7(6):374–387, 2007.

Schmeidler, D. Subjective probability and expected utility without additivity. *Econometrica: Journal of the Econometric Society*, 57(3):571–587, 1989.

Sims, C. A. Pitfalls of a minimax approach to model uncertainty. *American Economic Review*, 91(2):51–54, 2001.

Stein, C. et al. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. The Regents of the University of California, 1956.

Thimme, J. and Völkert, C. Ambiguity in the cross-section of expected returns: An empirical assessment. *Journal of Business & Economic Statistics*, 33(3):418–429, 2015.

Tütüncü, R. H. and Koenig, M. Robust asset allocation. *Annals of Operations Research*, 132(1-4):157–187, 2004.

Uppal, R. and Wang, T. Model misspecification and underdiversification. *The Journal of Finance*, 58(6):2465–2486, 2003.

Von Neumann, J. and Morgenstern, O. Theory of games and economic behavior, NJ, Princeton University Press Princeton, 1945.

Wald, A. Statistical decision functions. 1950.

Wang, Z. A shrinkage approach to model uncertainty and asset allocation. *Review of Financial Studies*, 18(2):673–705, 2005.

Zhang, J., Jin, Z., and An, Y. Dynamic portfolio optimization with ambiguity aversion. *Journal of Banking & Finance*, 79:95–109, 2017.

Junior Management Science 4(3) (2019) 339-391

Junior Management Science

journal homepage: www.jums.academy

Mobile App Service Quality Dimensions and Requirements for Mobile Shopping Companion Apps

Tobias Wulfert

Westphalian Wilhelms-University, Münster

Abstract

The increasing utilization of mobile apps for shopping leads retailers to provide customers with dedicated mobile shopping companion apps to create an omni-channel shopping experience involving traditional brick-and-mortar, electronic and mobile business. Mobile shopping companion apps extend the traditional and electronic services of brick-and-mortar retailers by an additional mobile channel providing the customer with a digital companion supporting the shopping within and outside the stores using mobile technology. A twofold approach is pursued in this thesis. Firstly, a structured literature review is conducted to identify candidate dimensions for developing a scale for measuring the service quality of mobile shopping companion apps. Secondly, design requirements for improving the service quality of these mobile apps are deduced from online customer reviews of three exemplary mobile shopping companion apps applying a qualitative content analysis. The mobile app service quality of mobile shopping companion apps can be measured using a hierarchical and multi-dimensional scale consisting of three primary dimensions, seven secondary dimensions and 22 related items. The primary dimensions interaction quality, environment quality and outcome quality structure the secondary dimensions responsiveness, information, security and privacy, design, performance, technical reliability and valence. Based on these dimensions, 22 implementation guidelines and 14 service design requirements are derived as potential areas for optimizing the mobile app service quality of mobile shopping companion apps and achieving a high overall service quality. A mobile shopping companion app should include a set of features consisting of 16 features from three different areas. Results show that measuring the service quality of mobile shopping companion apps require for a tailored measurement scale. Equally, design requirements are proposed for this particular category of mobile apps. Retailers should provide a single mobile shopping companion app providing all features and mobile services to the customer.

Keywords: mobile service; mobile commerce; shopping companion; service quality.

1. Introduction

The ubiquity of smartphones including the possibility to be always connected to the internet allows customers to buy products using mobile applications (apps) on the move without temporal or spatial constraints (Heinonen and Pura, 2008, p. 4). Innovations such as the introduction of the smartphone and the increasing speed of cellular networks impact the retail sector by converging the retailers' value chains (Grewal et al., 2017; Verhoef et al., 2015). Customers increasingly use mobile apps and provided mobile services (m- services) for shopping leading retailers to provide their customers with dedicated mobile apps for shopping to create an omni-channel shopping experience (Yang and Kim, 2012, p. 778).

Overall, the Mobile Commerce (m-Commerce) revenue and share of the retail sales is rapidly growing with a projected revenue of USD 156 billion for the American market for 2017 (EMarketer, 2017; Meola, 2016). While the traditional brick-and-mortar retailing and Electronic Commerce (e-Commerce) only grew by three respectively 14 percent in the first quarter of 2017, m-Commerce revenues increased by 43 percent (Lipsman, 2017, p. 37). This increase is mainly caused by high growth rates for smartphone-based mobile shopping (EMarketer, 2016). Further, a steadily increasing number of customers prefer to use the mobile channel for shopping while at the same time spending more money on m-Commerce (Meola, 2016). In the first quarter of 2017, nearly every fifth dollar was spent by the customers using mobile devices (Lipsman, 2017, p. 8).

Acknowledging the increasing importance of electronic

DOI: https://doi.org/10.5282/jums/v4i3pp339-391

and mobile sales and distribution channels as drivers for overall company success and future development and the stagnating role of the traditional brick-and-mortar business, traditional retailers previously only operating brick-and-mortar stores recently entered e-Commerce and m-Commerce providing additional electronic services (e-services) and m-services and especially mobile shopping companion apps for their customers (Lu et al., 2009; Kuo et al., 2016; Rudolph et al., 2015). Mobile shopping companion apps executed on mobile devices provide an omni-channel experience for the customers supporting the shopping within and outside the stores. They support and enrich the customers' online shopping experience and extend the features of general mobile shopping apps by accompanying the customer in-store enhancing the customers' offline shopping experience in the brick-and-mortar stores. Thus, they offer several features that are neither usefully realizable on stationary channels nor the traditional internet (Rudolph et al., 2015, p. 43).

With an increasing number of brick-and-mortar retailers also competing in the m-service environment with mobile apps and customers demanding higher quality services, the measurement of the retailers own Mobile App Service Quality (MASQ), enabling the specific improvement of their m-services is essential to those retailers (Huang et al., 2015; Kuo et al., 2016). Service Quality (SQ) is an important determinant for the success of a company influencing "business performance, lower cost, customer satisfaction, customer loyalty, and profitability" (Seth et al., 2005, p. 913). Thus, brick-and-mortar retailers aim at providing the best MASQ to further increase their mobile and traditional sales and to put themselves in a superior position compared to their competitors (e.g., Bauer et al., 2006; Kuo et al., 2016; Lu et al., 2009). For measuring the MASQ of the mobile shopping companion apps the "disconfirmation paradigm" (Brady and Cronin, 2001, p. 35) that is already used for the early conceptualizations of the SQ construct is transferred to the context of mobile apps (Grönroos, 1984; Parasuraman et al., 1985). It defines SQ as the difference between perceived and expected quality of a service (Brady and Cronin, 2001, p. 35).

The definition of mobile shopping companion apps as a shopping companion for in-store and mobile shopping demands for a tailored scale to assess the SQ of the specific features and m-service provided by the mobile shopping companion app in the m-Commerce environment (Ladhari, 2010, p. 474). Thus, existing scales for measuring Electronic Service Quality (ESQ) and Mobile Service Qualitys (MSQs) do not cover all aspects of the MASQ. Existing scales do not fully consider the impact of m-service delivery on in-store and mobile shopping simultaneously. Applying ESQ and MSQs measurement scales on mobile shopping companion apps may result in inconsistent results. Unfortunately, research on service provisioning via mobile shopping apps on mobile devices is very rare. One scale for measuring MASQ is provided by Kumar (2017) however it contains inconsistencies in the line of argumentation and is not published in a renowned journal. Another approach is presented by Kuo et al. (2016) in a conference paper without elaborating on the proposed measurement dimensions. This thesis contributes to the extant literature by introducing a dedicated measurement scale for assessing the SQ of mobile shopping companion apps and extensively elaborating on its dimensions.

> RQ1. How can the service quality of mobile shopping companion apps be assessed?

Most of the previous papers on ESQ and MSQ close after the development of a scale assessing the quality of a service (e.g., Parasuraman et al., 2005; Stiakakis and Petridis, 2014). Only a few deliver concrete suggestions on how to improve e-services and m-service with regards to the identified SQ dimensions (e.g., Knote and Söllner, 2017). Again, these suggested design requirements are only partly transferable to the specifics of mobile shopping companion apps. Besides being able to measure their MASQ, retailers demand for concrete design requirements as suggestions for improving their mobile shopping companion apps. As SQ is perceived by the retailers' customers, the users of the mobile apps, their perceptions and opinions have to be taken into account for deriving design requirements for each of the identified dimensions. Hence, a further contribution of this thesis is to derive design requirements from online customer reviews that might lead to a high SQs for mobile shopping companion apps according to the MASQ measurement scale. For this thesis, the online customer reviews for the mobile shopping companion apps provided by Walmart Stores, Inc. (Walmart), Tesco PLC (Tesco) and Marks and Spencer plc (M&S) are analyzed as examples for major mobile apps in the United States and the United Kingdom. Online customer reviews are a form of electronic word-of-mouth (eWOM) and are publicly accessible in the mobile app stores (Hennig-Thurau et al., 2004, p. 39). These design requirements should involve functional and non-functional requirements such as aesthetic or performance aspects.

> RQ2. What are design requirements for a high service quality when developing mobile shopping companion apps?

The main emphasis of this thesis is put on the derivation of design requirements for a high MASQ and an explicit empirical analysis of the applicability and validity of the various dimensions in the context of the MASQ measurement scale is out of focus. The applicability and validity of the used MASQ dimensions is already proven in the context of ESQ and MSQ and implicitly validated within the Qualitative Content Analysis (QCA) identifying relevant online customer reviews for each of the MASQ dimensions. To answer these research questions, the following course of investigation is undertaken that follows the content structuring approach for a QCA (Mayring, 2015, p. 98):

Following this general introduction, the main constructs underlying this thesis are introduced and defined in the theoretical background chapter. Mainly, these are the development of SQ, the delimitation of mobile shopping companion

apps from general mobile apps and the introduction of online customer reviews. Then, the twofold approach applied in this thesis consisting of a Structured Literature Review (SLR) for searching previous literature on ESQ and MSQ as basis for the MASQ dimensions and a QCA for analyzing the online customer reviews in a structured manner is presented (Mayring, 2015; Vom Brocke et al., 2009). The scale for measuring MASQ is established in the next chapter based on the results of the SLR. After the explanation of the scale development and its presentation, the theory-based definition of the single dimensions is derived. The structure of this chapter follows the primary and related secondary dimensions of the MASQ scale. Next, design requirements for the developed MASQ dimensions are derived from the online customer reviews. The presentation of the retailers and their analyzed mobile shopping companion apps is followed by an overview of the sample of online customer reviews exported from the Mobile Application Distribution Platforms (MADPs). The illustration of the design requirements for high quality mobile shopping companion apps also follows the primary and secondary dimensions for measuring the MASQ identified in the previous chapter. The chapter closes with an analysis of features required for mobile shopping companion apps. Afterwards, the resulting MASQ measurement scale and the design requirements and their derivation are discussed incorporating underlying assumptions of this thesis and open topics. This thesis closes with a short summary and an outlook on possible future research directions.

2. Theoretical Background

The following chapter introduces definitions for the main terms and basic constructs that build the foundation for the rest of this thesis. Firstly, a general introduction to design requirements, quality in general and services is provided. Secondly, the development from SQ towards MASQ is shown, followed by a delimitation of mobile shopping companion apps from general mobile apps. Then, online customer reviews in the context of mobile app stores are narrowed down.

2.1. Design Requirements, Service and Quality

Design requirements describe "what the system should do" (Sommerville, 2011, p. 83) including provided services and potential operational constraints. The requirements include essential conditions or capabilities "needed by a user to solve a problem or achieve an objective" (Pohl, 2010, p. 16). For the application developers, the design requirements are often summarized in guiding design principles (Stahlknecht and Hasenkamp, 2005, p. 246). The term requirement stems from requirements engineering in the context of software development. Requirements can be differentiated in functional and qualitative requirements and constraints (Pohl, 2010, p. 17). Functional requirements are concerned with the main functions that describe the scope of a service including in- and outputs (Stahlknecht and Hasenkamp, 2005, p. 245). Quality requirements define quality properties "of the entire system or of a system component, service, or function" (Pohl, 2010, p. 18). Wiegers and Beatty (2013) summarize the quality requirements in eight requirements important for users and four requirements important for developers (p. 197). "A constraint is an organizational or technological requirement" (Pohl, 2010, p. 22) restricting the development process which deals among other things with system realization, quality, usability, reliability, safety and costs (Stahlknecht and Hasenkamp, 2005, p. 245). Design principles in general support the software design (Heinrich et al., 2004, p. 605). A principle in general guides the thinking, acting and behavior and is often used in practice if standardized methods are not applicable in a given context (Heinrich et al., 2004, p. 507). Within this thesis design requirements subsume requirements for the design of mobile shopping companion apps and the design of the m-services they provide.

A service "is defined as the co-creation of value between service systems (customers, providers, etc.), and service systems resources (the dynamic configuration of people, technology, organizations, and shared information) connected internally and externally by value propositions" (Hsu, 2009, p. 7). Norman (1984) describes services as "social acts which take place in direct contact between the customer and representatives of the service company" (p. 14). Services should provide value to the recipient of a service (Heinrich et al., 2004, p. 196). The service provider possesses the equipment and know-how for the execution of the service and typically charges the recipient with a service fee (Alpar et al., 2011, p. 81). Services can be characterized as heterogeneous, immaterial and intangible. Further, service production and consumption are inseparable and happen simultaneously (Grönroos, 1984; Sower et al., 2001). In the retailing context, services are typically "rendered to the consumer" (Grönroos, 1984, p. 38). A single service encounter describes “a period of time during which a consumer directly interacts with a service” (Bitner, 1990, p. 72). It can take place without interpersonal interaction in the e-service or m-service environment.

Quality is defined by the International Organization for Standardization (2008) as the "degree to which a set of inherent characteristics fulfills requirements". However, a universal definition of quality does not exist in literature. Distinct definitions of quality exist within nearly each research domain. So, for each quality related research it has to be decided which quality construct to choose (Reeves and Bednar, 1994). The authors define quality along excellence, value, conformance to specifications, and meeting or exceeding expectations. In the context of services, quality is often defined from a customer-based perspective as "meeting or exceeding customer expectations" (Evans and Lindsay, 2011, p. 9).

2.2. Development towards Mobile App Service Quality

Historically, SQ measures the quality of interpersonal, non-internet-based service interactions between a customer and a company (Li and Leng, 2008, p. 1). First conceptualizations are introduced in the 1980s and 1990s (Cronin and Taylor, 1994; Grönroos, 1984; Parasuraman et al., 1985).

With the commercialization and dissemination of the internet since 1994 (Fischer and Hofer, 2008, p. 414f.), services could be accessed electronically through websites even without the presence of a service provider's employee. The omission of the interpersonal interaction requires for adapted and new dimensions to measure the ESQ. Following, the development and increasing downstream of mobile cellular networks with the launch of the universal mobile telecommunications system in 2004 and introduction of mobile internet devices such as the Personal Digital Assistant (PDA) and smartphone, enables customers to access the provided services without timely and spatial constraints (Rudolph et al., 2015; Stahlknecht and Hasenkamp, 2005). Again, this change in the service provision propels the need for new scales to measure the MSQ. As smartphones further evolve and enable the execution of dedicated, third-party mobile apps (Rudolph et al., 2015, p. 43) provided via MADPs since 2008 (Genc-Nayebi and Abran, 2017, p. 208), this new channel of service provisioning demands new aspects to be considered when measuring the MASQ. This development is illustrated in figure 1 and underpinned with further details in the following subsections.

2.2.1. Service Quality

Because of the service characteristics mentioned in section 2.1 the objective measurement of the SQ is more difficult than the measurement of product quality (Sower et al., 2001). Research on SQ strives to establish frameworks and dimensions for SQ measurement. SQ theory is grounded "in the product quality and customer satisfaction literature" (Brady and Cronin, 2001, p. 34). There exist two early conceptualizations of the SQ construct by Grönroos (1984) called "Nordic" (Brady and Cronin, 2001, p. 35) model in literature and by Parasuraman et al. (1985) also known as "gap model" (van Dyke et al., 1997, p. 197). Both SQ constructs make use of the disconfirmation paradigm suggesting that quality is the result of the comparison between perceived and expected performance (Brady and Cronin, 2001, p. 35). In the context of SQ, perception relates to "the consumer's beliefs concerning the received or experienced service" (Jun et al., 2004, p. 819). Expectations depict a customer's desires or wants about the provided service (Parasuraman et al., 1988, p.17). Implicitly, this means that perceived SQ is a subjective matter, individual to each single customer (Rust and Oliver, 1994). Thus, SQ is defined as the degree of "discrepancy between customers' expectations and perceptions" (Parasuraman et al., 1994, p. 111) towards a received or experienced service (Parasuraman et al., 1985, 1988; Zeithaml et al., 2000). Perceived SQ "is a global judgment, or attitude, relating to the superiority of the service" (Parasuraman et al., 1988, p. 16), compared to customer satisfaction that describes "the result of specific service transactions" (Jun et al., 2004, p. 822). High SQ is achieved if the discrepancy between expectations and perceptions is marginal or if the perceptions exceed the expectations (Parasuraman et al., 1985).

As first SQ dimensions Grönroos (1984) identified technical and functional quality. While the former describes the outcome of the service encounter that is received by the customer, the latter indicates the quality of the service delivery, the transfer of the technical outcome to the customer, itself. Parasuraman et al. (1988) reduce their initial ten dimensions of SQ to the SERVQUAL scale consisting of the five dimensions tangibles, reliability, responsiveness, assurance and empathy (p. 23; fig. 2). Tangibles comprise the quality assessment of "physical facilities, equipment and appearance of personnel" (Parasuraman et al., 1988, p. 23). The accurate performance of the service by the service provider is summarized in the reliability dimension and responsiveness is considered with customer support. Further, assurance includes the ability of employees to be trustful and knowledgeable. The last dimension, empathy, involves customer care and individualized responses to customer inquiries (Parasuraman et al., 1988, p.23). The gap model by Parasuraman et al. (1985) incorporating the SERVQUAL scale is shown in figure 2.

It is of interest for companies to measure and improve the quality of their services because a high SQ positively influences "business performance, lower costs, customer satisfaction, customer loyalty and profitability" (Seth et al., 2005, p. 913). The positive effect on the customer's satisfaction with the used service is highlighted in several papers (e.g. Jun et al. 2004; Lee and Lin, 2005).

2.2.2. Electronic Service Quality

Analogous to traditional service environments, quality is an important determinant for the success of e-services (Santos, 2003; Zeithaml, 2002). A high ESQ is even more important than offering low prices on a web presence (Zeithaml, 2002). Furthermore, it has positive effects on both customer satisfaction and customer web site loyalty (Cristobal et al., 2007; Gefen, 2002; Swaid and Wigand, 2009). Thus, ESQ highly influences the retailer's e-Commerce sales and revenue and even its strategy (Santos, 2003).

Although some authors found evidence for the applicability of SERVQUAL in e-service environments (Pitt et al., 1995), the majority of the literature argues that e-services differ too much from traditional services because of the absence of a physical, interpersonal interaction that is mostly substituted by the interaction with the online retailer's website (Jun et al., 2004) so new scales for ESQ assessment need to be established (e.g., Cox and Dale, 2001; van Dyke et al., 1997; Kettinger and Lee, 1997). E-services differ from traditional services across several characteristics as the services are provided using electronic networks (Jun et al., 2004; Rust and Lemon, 2001). Firstly, the cost structure consists of high fixed costs for the development and only marginal costs for the service provisioning. Secondly, a high degree of technical standardization together with huge parts of the service provisioning in back offices fosters a high degree of outsourcing. Thirdly, new e-services can be rapidly developed and continuously improved through updates. Finally, customer interaction with the e-service environment can be tracked to allow for transparent service feedback (Riedl et al., 2009).

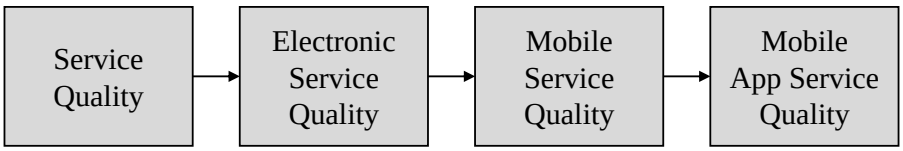


Figure 1: Evolution of Service Quality

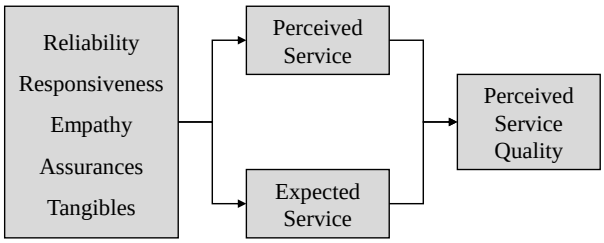


Figure 2: Gap-Model (Parasuraman et al., 1988)

Parasuraman et al. (2005) define ESQ as "the extent to which a web site facilitates efficient and effective shopping, purchasing, and delivery of products and services" (p. 217). Thereby, it “includes both pre- and post-Web site service aspects” (Zeithaml, 2002, p. 136) such as information retrieval, payment and product or service obtainment (Collier and Bienstock, 2006; Francis and Lesley, 2002; Parasuraman et al., 2005; Santos, 2003). Many of the existing scales for measuring ESQ do not cover the whole purchasing process (Blut et al., 2015). Thus, they can be applied by only browsing through a given website (Francis, 2009). ESQ goes beyond the mere assessment of the design of the service provider’s website as some authors suggest measuring the Website Design Quality (WDQ) to determine the ESQ (e.g., Yoo and Donthu, 2001; Aladwani, 2002; Barnes and Vidgen, 2002).

As the ESQ encompasses all phases of a customer’s interactions with a web site (e.g., Li and Leng, 2008; Swaid and Wigand, 2009; Wolfinbarger and Gilly, 2003), Bauer et al. (2006) base their ESQ dimensions upon a four-stages e-Commerce process consisting of an information phase, agreement phase, fulfillment phase and after-sales phase (fig. 3). As customers are likely to evaluate the quality of the overall shopping process and process outcome instead of single phases (Kar et al., 2003), ESQ scales should cover the whole shopping process from the first visit of the website to the after-sales phase.

Even though Zeithaml et al. (2000) state that the evaluation of ESQ does not require to consider the type of the product or service involved, more recent research distinguishes between "pure" e-services dealing with non-tangible products or services and "non-pure" e-services (Vlachos and Vrechopoulos, 2008, p. 5) dealing with tangible products in the e-commerce context. For tangible products, the item availability is important for the service assessment (Parasuraman et al., 2005) while instant availability is usually always the case for non-tangible products such as software or music downloads (Vlachos and Vrechopoulos, 2008).

2.2.3. Mobile and Mobile App Service Quality

Prior research has proven that SQ is also an important factor for m-services. It positively affects the customer satisfaction and loyalty with the m-service provider and thus is a determinant for the provider’s success (Kuo et al., 2016, p. 2). Especially in the context of m-Commerce, SQ measurement is vital for improving the retailer’s m-services (Lu et al., 2009, p. 229). High quality m-services can be a future competitive edge for a retailer (Grewal et al., 2017, p. 2).

M-services differ a lot from traditional and e-services and need to be investigated separately because of certain specific characteristics (Stiakakis et al., 2013, p. 1). Similar to e-services, a personal interaction between the customer and a retailer’s employee will not take place. The interface for accessing the m-service will be a mobile device or more specifically a mobile app (Zhao et al., 2012). However, the main characteristics of m- services is the absence of temporal and spatial constraints (Heinonen and Pura, 2008, p. 4). These services "can be accessed on the move, where and whenever the need arises" (Heinonen and Pura, 2008, p. 2). When designing m-services time-sensitivity, location- awareness and personalization have to be taken into account. Time-sensitivity refers to the currentness of information while location-awareness highlights the ability of m- services to incorporate location-based information in the service delivery. As mobile devices such as smartphones are omnipresent to the user, personalizing the m-service to the user’s requirements is increasingly important (Lee and Benbasat, 2004). Clarke and Flaherty (2003) summarize the benefits of m-services in the factors ubiquity, convenience, localization and personalization differentiating m-services from e-services. Because of these characteristics, m-services can be used throughout the "traditionally distinct spheres of life" (Rao and Troshani, 2007, p. 62) in a seamless and pervasive manner. Therefore, Stiakakis et al. (2013) propose

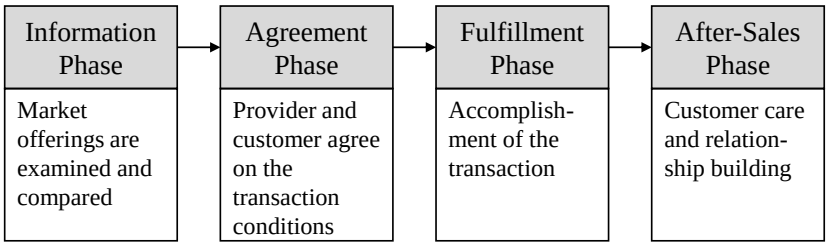


Figure 3: Stages in the e-Commerce Process (Bauer et al., 2006)

to identify quality dimensions specifically for m-services not copying dimensions from ESQ.

The concept of mobility is not well established in literature. However, it can be described as "moving around - either in time or in space" (Bouwman et al., 2012, p. 67). M-services are defined as activities "that occur when mobile consumers interact with systems or service provider employees with the support of a mobile telecommunications network" (Rao and Troshani, 2007, p. 62). Hence, MSQ measures the perceived SQ of m-services (Sti- akakis and Petridis 2014). The literature on MASQ does not provide a dedicated definition for the SQ of mobile apps (Kumar, 2017). Nevertheless, a differentiation between MSQ and MASQ can be made regarding the service provisioning. While MSQ involves any m-service accessible via a mobile device such as mobile internet, location-based services, voice transmission or short and multimedia message service (Bouwman et al., 2012, p. 68), MASQ is more concerned with the mobile apps executed on the mobile device providing the m-services to the users. Thus, MASQ can be seen as a subset of MSQ.

2.3. Mobile Shopping Companion Apps

Before the classification and definition of mobile shopping companion apps a short introduction to its m-Commerce context is given.

2.3.1. Mobile Commerce as Context

Within the last years retailers extended their traditional in-store business by an additional online channel (Rudolph et al., 2015). These retailers are called "multi-channel retailers" (Xing et al., 2010, p. 417), "bricks and clicks" (Xing et al., 2010, p. 417) or "Clicks-and- Mortar" (Laudon et al., 2010, p. 615). In contrast to them "pure online retailers" (Xing et al., 2010, p. 417) do not operate any up-front stores and their only sales channel is the internet. Multi-channel retailers apply "The Retail Model" (Clark, 1997, p. 17) offering their retail products online and extend the online shop with additional services such as store finders or in-store product availability monitoring. Following the recent developments of mobiles devices and the mobile internet, retailers provide their online services also on mobile devices and implemented dedicated mobile apps (Huang et al., 2015, p. 126). These shopping apps do not only make the e-services accessible on mobile devices but also add new features making use of the characteristics of m-services already mentioned above. Compared to mobile websites, mobile shopping apps offer improved security features and keep the customers in the retailer's self-contained environment (Taylor and Levin, 2014). These mobile shopping apps allow the retailers to create a "seamless omni-channel experience" (Rigby, 2011, p. 68) and simplify the shopping process by supporting the customer (Rudolph et al., 2015).

The importance of m-Commerce is increasing and can be measured by the revenue generated (EMarketer, 2017) and the overall time spent with shopping apps (Solomon, 2015). In m-Commerce "transactions with monetary value" (Clarke, 2001, p. 41) are executed "focusing on the trading of goods and services" (Zobel, 2001, p. 3) involving at least one of the business partners using a wireless connection (e.g. cellular network or Wireless Lo- cal Area Network (WLAN)) with a mobile device such as a smartphone, tablet computer or PDA (Laudon et al., 2010; Silberer, 2002; Stahlknecht and Hasenkamp, 2005). At the same time m-Commerce is a subset of all e-Commerce activities and a discrete category involving new technologies that are able to substitute e-Commerce activities and enabling completely new m-Commerce services that are not possible in a wired e-Commerce environment (Zobel, 2001). Besides the possibility to simply access websites with a mobile device, smartphones allow to execute m-Commerce transactions by installing mobile apps deployed by the retailers and other third parties. A smartphone is a mobile device combining speech and textual functions of a mobile phone with data transmission and app execution capabilities of a PDA (Laudon et al., 2010, p. 340).

2.3.2. Classification and Definition

Generally, Nickerson et al. (2009) "define a mobile application as a use of a mobile technology by an end-user for a particular purpose" (p. 7). In case of this definition the purpose can be anything from entertainment to making an airline reservation or doing the shopping. Although a categorization of mobile apps is difficult due to the large number of diverse mobile apps and a limited amount of textual information about each of the apps (Zhu et al., 2014), Put et al. (2014) describe mobile shopping applications as mobile apps that enhance the customers' shopping experience and that are usually executed on a smartphone. These mobile apps are provided by a retailer and installed by a customer on his or her mobile device (Rudolph et al., 2015, p. 43). Indeed,

this is still a broad definition of mobile shopping apps.

Before giving a definition for mobile shopping companion apps, they are classified applying the taxonomy of mobile apps including seven dimensions introduced by Nickerson et al. (2009). Regarding the temporal dimension is the interaction between the user and the application synchronous. The user is informed about the current stock and price information in real time. The direction of information flow as described in the communication dimension is interactional meaning that the mobile app provides the customer for instance with product information and the customer provides information on the products he or she wants to purchase. As purchase transactions including mobile payments can be executed within the mobile app, the transaction dimension for mobile shopping companion apps is transactional. The target group of mobile shopping companion apps is the general public allowing every individual who is interested in using the mobile app to download it. Therefore, the public dimension is public. Concerning the multiplicity dimension each user has the impression that he or she is individually using the mobile app. Group interaction is only partially visible if other users' product reviews are publicly visible. The location dimension introduces the context-awareness to the taxonomy. Mobile shopping companion apps are location-aware as they can provide the user with the location of the nearest store and store related advertising based on the user's location. From the identity dimension point of view, the apps are identity-based. They incorporate customer accounts to enable online shopping, integrate premium programs and their appearance is sometimes customizable to the users' preferences up to a certain extent.

Mobile shopping companion apps can be defined as mobile shopping apps extending the traditional and e-services of brick-and-mortar retailers with an additional mobile channel providing the customer with a digital companion supporting the shopping within and outside the stores using mobile technology. As mobile shopping companion apps make use of m-services that provide the infrastructure for the deployment of the mobile app (Nickerson et al., 2009), mobile shopping companion apps can be further defined by describing the m-services they provide to the user. Using the classification of mobile apps proposed by Varshney and Vetter (2002), mobile shopping companion apps incorporate mobile financial services, mobile advertising services, product locating services and shopping services.

While mobile shopping is often considered as a sole online distribution channel (Groß, 2015), mobile shopping companion apps provide several additional services besides the interface to the retailer's online shop. Following Gummerus and Pihlström (2011), Rudolph et al. (2015) identified the most attractive services of mobile shopping apps using an online survey among the dimensions information, esteem, monetary, convenience and situational value (p. 45). Mobile shopping companion apps contain a barcode scanner for retrieving detailed product information and prices, a tab for managing the retailer's loyalty cards and allow for in-store and online payment using the mobile app. Furthermore, they allow for monitoring the order progress and create shopping lists in addition to notifications for special offers and in-store sales, a tool to analyze sales receipts and offer shopping proposals based on previous shopping. For locating the nearest store location, they make use of the location-awareness. Real-time information enables to check the current in-store and online product availability. Finally, the mobile apps allow information sharing on social media sites and read product reviews created by other customers (tab. 1). As the mobile shopping companion apps are implemented by retailers traditionally Value dimension Services provided by mobile shopping companion apps Monetary Barcode scanner, loyalty card manager, in-app payment Convenience Online purchase, order progress monitoring, shopping lists Information Notifications, sales receipts analyzer, shopping proposals Situational Store locator, in-store and online product availability Social Products reviews, social media link operating brick-and-mortar stores selling nearly exclusively tangible products, this is a form of non-pure m-Commerce. Rudolph et al. (2015) found that mobile shopping companion apps are mostly used in the information and agreement phase (fig. 3) close to the purchase. Mobile shopping companion apps are an additional communication channel in the dyadic constellation of a retailer and its customers (fig. 4).

2.4. Customer Reviews for Mobile Applications

MADP[1] incorporate an opinion platform to allow the users to provide feedback for downloaded and used mobile apps. A MADP, also called mobile app store, is a type of digital distribution that makes mobile apps available for download for smartphone users without the need for physical distribution media (Dibia and Wagner, 2015, p. 4304). In 2008 the first MADP was launched by Apple (Genc-Nayebi and Abran, 2017, p. 208). Until the start of the Mac App Store and the Windows Store for desktop computers and laptops (Idu et al., 2011; Microsoft, 2017), the term app store was solely linked with MADPs.

The feedback published in the MADPs can be characterized as eWOM that is defined as "any positive or negative statement made by potential, actual, or former customers about a product or company, which is made available to a multitude of people and institutions via the Internet" (Hennig-Thurau et al., 2004, p. 39). In case of this thesis, eWOM provided through the Apple App Store (App Store) and Google Play Store (Play Store) is analyzed. It can take the form of ratings or customer review messages (Trenz and Berger, 2013, p. 3). Both eWOM mechanisms are publicly available and visible for other users and the developers of the mobile app (Pagano and Maalej, 2013, p. 125). Both, the Apple App Store (App Store) and Play Store include a rating system on a one to five scale symbolized by stars (Fu et al., 2013, p. 1278). While the App Store only allows half

[1] To avoid mix-ups between Apples App Store and the generic term app store, for the rest of this thesis the term Mobile Application Distribution Platform (MADP) is used.

Table 1: Services Provided by Mobile Shopping Companion App

Value dimension	Services provided by mobile shopping companion apps
Monetary	Barcode scanner, loyalty card manager, in-app payment
Convenience	Online purchase, order progress monitoring, shopping lists
Information	Notifications, sales receipts analyzer, shopping proposals
Situational	Store locator, in-store and online product availability
Social	Products reviews, social media link

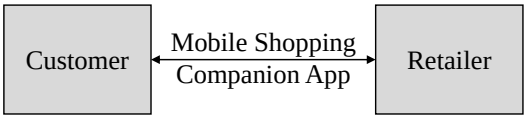


Figure 4: Dyadic Perspective on Mobile Shopping Companion Apps

star intermediate steps, the Play Store shows the actual mean value of all ratings.

Online customer reviews are defined as "peer-generated product evaluations posted on company or third party websites" (Mudambi and Schuff, 2010, p. 186) that are influenced by personal and subjective experiences of a product or service (Trenz and Berger, 2013, p. 2). In case of this thesis, the product resembles the mobile shopping companion apps and a MADP is used for posting the reviews. Users tend to write reviews either if they are euphoric and positively impressed by the use of the app or disappointed about malfunctions or the mobile app in general. Additionally, these reviews are often written using mobile devices on which typing is more difficult and not so much attention is paid on formatting and writing style (Fu et al., 2013, p. 1276). Customer reviews tend to be very short in nature and vary strongly in their quality relating to content and expressiveness (Fu et al., 2013; Genc-Nayebi and Abran, 2017; Pagano and Maalej, 2013) The online customer reviews can be used for "user-driven quality assessment" (Pagano and Maalej, 2013, p. 1) and help to improve the mobile apps or to add new functions desired by the users. Reviews are intended to inform other potential users about the usability and usefulness of the app and to guide the further development. Moreover, PPagano and Maalej (2013) have found that the topics of customer reviews for mobile apps in MADPs cover at least one of the themes "community", "requirements", "rating" and "user experience" (p. 130).

3. Methodology

Within this thesis a twofold approach is applied using two methodologies taken from literature to address the raised research questions. Firstly, a SLR is conducted to develop a multi-dimensional and hierarchical scale to measure the SQ of mobile shopping companion apps based on prior findings in literature (Vom Brocke et al., 2009). Secondly, online customer reviews of three distinct mobile shopping companion apps are analyzed in a QCA to derive design requirements for each dimension of the scale (Mayring, 2015).

3.1. Structured Literature Review

To identify and evaluate appropriate literature for this thesis the SLR approach presented by Vom Brocke et al. (2009) is applied not explicitly on Information Systems (IS) literature but on literature related to service and quality research and retailing literature especially in the field of e-Commerce and m-Commerce. The author's iterative review process consists of the four main phases definition of review scope, conceptualization of topic, literature search and literature analysis and synthesis.

The review scope of the present thesis is defined using the taxonomy of literature reviews by Cooper (1988). The SLR integrates and synthesizes previous research outcomes and theories on a conceptual basis in the context of ESQ, MSQ and MASQ (app. A). The conceptualization of the topic phase has already been addressed through the definitions in chapter 2. The next phase, literature search, itself consists of the sub phases journal search, database search, in-/ exclusion criteria and back-/ forward search. Concerning the searched journals, no restrictions were made during the SLR to allow for an exhaustive coverage. The topic of SQ spreads into various research areas (app. A). Four online databases were searched to identify appropriate literature. EBSCOhost, Scopus and Web of Science reference to papers from the majority of renowned peer-reviewed, scientific journals including retailing, service and quality related articles. The electronic library of the Association for Information Systems (AIS) is included to integrate research papers, conference papers and journal articles from the field of IS research. As Vom Brocke et al. (2009) stated that a precise set of keywords should be used to find appropriate literature in online databases, prior

to the SLR pivotal literature to the topic is checked for keywords used that are in line with the research questions of the present thesis. Structuring those keywords results in the general search string for the online database search as depicted in table 2 (app. A).

The number of initial findings from each online database can be seen in table 3. In- and exclusion criteria have been specified to ensure that only papers relevant for the backdrop of the research questions are selected. Papers are included that focus on the development and evaluation of scales for the measurement of ESQ, MSQ and MASQ. The complete set of exclusion criteria is presented in appendix A. Most importantly papers are excluded whose SQ dimensions are not applicable generally or in the context of e-Commerce and whose focus of the SQ dimensions is not Business to Consumer (B2C). Further, papers are excluded that are not available in English language, duplicates, including papers repeating previously established SQ dimensions. Applying these criteria on the initial set of papers results in 12 relevant papers (tab. 3).

After the set of relevant papers has been selected an additional backward search is conducted with the references of the selected papers and a forward search is executed to find further paper that cited the already selected papers. This results in additional 23 potentially relevant papers. Evaluating these additional papers combined with the already relevant papers results in the final set of 35 relevant papers. These papers are presented in table 4 including quality and measurement domain as well as the number of dimensions[2] and items. The SLR reveals that some of the ESQ scales do not measure the quality of the whole e-Commerce process but only focus on the design of the website. Thus, they are labeled WDQ scales (tab. 4) in line with the findings by Cristobal et al. (2007) and Francis (2009). The final set of papers forms the foundation for the literature analysis and synthesis in chapter 4. This synthesis results in a scale to measure the SQ of mobile shopping companion apps. The dimensions for this scale are derived by comparing prior ESQ, MSQ and MASQ scales in a conceptual matrix using the dimensions by Stiakakis and Petridis (2014) as reference (Webster and Watson, 2002).

3.2. Online Customer Review Analysis

The customer reviews of the three mobile shopping companion apps[3] by Walmart, Tesco and M&S are analyzed to derive design requirements for the improvement of the MASQ. For the analysis of the online customer reviews the systematic approach for a QCA by Mayring (2015) is applied combined with the approach used by PPagano and Maalej (2013) and Knote and Söllner (2017) to extract and prepare the sample of reviews for the purpose of analysis. The QCA by Mayring (2015) is frequently used in literature to derive information from user-generated content like customer reviews (e.g., Balauca and Mehta, 2015; Giannakos, 2014; Spreer and Rauschnabel, 2016).

Following the two-step approach introduced by PPagano and Maalej (2013), in the preparation phase the customer reviews for the three mobile shopping companion apps are exported from the App Store and Play Store using a web-based customer review crawler service[4] on the 3rd of July 2017 (p. 127). Next, the exported customer reviews are converted from the comma-separated value format into a table-based Microsoft Excel file. These raw files contain information concerning the mobile app name, date, title and content of the review, the nick name of the customer who created the review, the rating that is provided and the rated version of the mobile app. Additional columns for the integration of the coding are added to the raw files (digital appendix).

In the first step of the analysis phase, the customer reviews are limited to current reviews from 2017 and useless and non-informational customer reviews with regard to the design requirements are discarded manually (Knote and Söllner, 2017, p. 125). Among others, these are reviews expressing praise like "lovely app" or "Great app!" and criticism without implications for improvements such as "Useless!" or "Can't recommend the app". Additionally, customer reviews concerning the retailer's general assortment, delivery quality, time and costs and recommendations for other users are excluded just like non-English reviews. This step leaves 1,448 reviews for the further QCA (tab. 13).

For the assignment of the customer reviews to the identified dimensions from the literature review, the QCA is applied (Mayring, 2015, p. 50ff.). More specifically, the content structuring approach a form of deductive category assignment (Mayring, 2014, p. 95f.) that allows the extraction of information according to theoretically predefined categories is used (fig. 5). The manual coding of the online customer reviews is supported by an online tool[5]. After the manual coding, it allows to export statistics on the coding categories including exemplary text passages (digital appendix).

The unit of analysis, the customer reviews for mobile apps, has already been defined in section 2.4 including the analysis of the situation of creation and the formal characterization of the material. Initially, 10,099 customer reviews are extracted from a total of 525,813 ratings. For further statistics on the extracted customer reviews see subsection 5.1.2.

The content categories resemble the MASQ dimensions and the characteristics equal the items as manifestations of the categories. These steps are the output of the literature analysis and synthesis in chapter 4. Moreover, definitions, anchor examples and the coding system are also included in this chapter.

[2] If the SQ scale uses a hierarchical structure, the secondary dimensions are presented in brackets.

[3] More information about the three mobile shopping companion apps by Walmart, Tesco and M&Sunder investigation in this thesis can be found in subsection 5.1.1 and appendix B.

[4] The used web-based customer review crawler service is provided by heedzy.com. To exceed the limit of freely extracting 100 customer reviews, a premium license is required.

[5] The online tool supporting the manual coding is QCAmap.org provided by Prof. Dr. Philipp Mayring and Dr. Thomas Fenzl.

Table 2: General Search String

*(("app" OR "mobile" OR "electronic" OR "m-" OR "e-") AND ("service quality") AND ("*commerce" OR "*shopping") AND ("criteria" OR "dimension" OR "measure*"))*

Table 3: Structured Literature Review Summary

	AIS Electronic Library	Web of Science	Scopus	EBSCO Host
Preliminary results	26	122	99	95
Title check and availability	11	41	48	33
Duplicate check		109		
Abstract check		42		
Content check		12		
Forward/Backward search		23		
Final results		35		

As Mayring (2015) suggests evaluating the dimensions and their items with a smaller test sample taken from the set of customer reviews, 100 customer reviews per mobile shopping companion app are analyzed to check the coverage of the identified MASQ dimensions (p. 97). The test run is indicated with the back loop and activity six in figure 5. This test run reveals that the dimension customization/personalization seems not to be of interest for the customers although it is often considered as a MASQ dimension and the mobile shopping companion apps allow at least for minor customization.

The analysis of the customer reviews, the paraphrasing and summary per category are all part of chapter 5 resulting in design requirements for an improved SQ for mobile shopping companion apps. Besides the anchor examples for the dimensions, further quotes taken from the online customer reviews will be integrated in chapter 5 following the prototyping approach (Mayring, 2015, p. 48). These prototypes should illustrate the derivation of the design requirements from the online customer reviews.

4. Mobile App Service Quality

This chapter is dedicated to the development and presentation of the MASQ scale. Firstly, the scale is introduced as multi-dimensional and hierarchical including the process of the scale development (section 4.1). Secondly, the primary dimensions structure the description of the secondary dimensions of the MASQ scale (sections 4.2-4.4).

4.1. Mobile App Service Quality Measurement Scale

As Ladhari (2010) mentions that SQ scales have to be tailored to the specific industry applied (p. 474), the developed measurement scale can be used to assess the MASQ of mobile shopping companion apps provided by brick-and-mortar retailers.

4.1.1. Dimensionality and Hierarchy

The SLR reveals that there is only little previous research available on MASQ. This may be because of m-Commerce increasingly emerged within the last few years making use of recent mobile technologies (Kuo et al., 2016, p. 2). Only the paper by Kuo et al. (2016) deals with MASQ (tab. 4). An additional attempt to develop a scale for the measurement of MASQ was made by Kumar (2017). However, this research was discarded from the final set of papers because of quality reasons. Therefore, the scale for measuring MASQ has to draw on existing MSQ scales as MASQ is described as a subset of MSQ. Additionally, some dimensions from ESQ and even WDQ can be adapted for measuring the SQ of mobile apps.

Though, there seems to be no consensus on the exact number and nature of dimensions and items when measuring ESQ or MSQ (Holloway and Beatty, 2008, p. 349). The number of dimensions reaches from two to 15 while the number of related items varies from eleven to 40 (tab. 4). Previous literature has already proven this finding for available ESQ scales (e.g., Blut et al., 2015; Ladhari, 2010). It is justified by diverse areas of application because of different focuses of the scales and particular industry segments (Hsu et al., 2012, p. 242). Nevertheless, it is broadly recognized in literature that SQ is measured using multi-dimensional scales (e.g., Collier and Bienstock, 2006; Cristobal et al., 2007; Heinonen and Pura, 2008; Wolfinbarger and Gilly, 2002). This is due to the fact that the quality and SQ constructs are complex (Collier and Bienstock, 2006; Parasuraman et al., 1985, 1988) as they encapsulate functional and non-functional aspects (Francis, 2009, p. 451). This is also the case for SQ in the electronic and mobile environment (Aladwani, 2002; Francis, 2009, p.

Table 4: Structured Literature Review Summary

Author	Year	Quality domain	Measurement domain	Number of dimensions/items
Yoo and Donthu	2001	WDQ	e-Commerce	4/9
Aladwani	2002	WDQ	General	6/25
Barnes and Vidgen	2002	WDQ	e-Commerce	5/25
Francis and Lesley	2002	ESQ	e-Commerce	6/23
Gefen	2002	ESQ	e-Commerce	3/22
Janda et al.	2002	ESQ	e-Commerce	5/22
Madu and Madu	2002	ESQ	e-Commerce	15/n.a.
Wolfinbarger and Gilly	2002	ESQ	e-Commerce	4/14
Cai and Jun	2003	ESQ	e-Commerce	4/19
Santos	2003	ESQ	e-Commerce	2(11)/n.a.
Wolfinbarger and Gilly	2003	ESQ	e-Commerce	4/14
Jun et al.	2004	ESQ	e-Commerce	6/21
Long and McMellon	2004	ESQ	e-Commerce	5/19
Webb and Webb	2004	WDQ	e-Commerce	2(9)/43
Cao et al.	2005	WDQ	e-Commerce	4(8)/42
Kettinger and Lee	2005	ESQ	General	4/18
Lee and Lin	2005	ESQ	General	5/15
Parasuraman et al.	2005	ESQ	e-Commerce	4/22
Parasuraman et al.	2005	ESQ	e-Commerce	3/11
Bauer et al.	2006	ESQ	General	5/25
Collier and Bienstock	2006	ESQ	e-Commerce	3(11)/54
Fassnacht and Koese	2006	ESQ	General	3(9)/24
Bressolles et al.	2007	ESQ	e-Commerce	6/18
Cristobal et al.	2007	ESQ	General	4/17
Loiacono et al.	2007	WDQ	General	4(12)/36
Holloway and Beatty	2008	ESQ	e-Commerce	4/37
Li and Leng	2008	ESQ	General	3(9)/n.a.
Tan and Chou	2008	MSQ	General	7/36
Vlachos and Vrechopoulos	2008	MSQ	General	7/32
Swaid and Wigand	2009	WDQ	e-Commerce	7/28
Ding et al.	2011	ESQ	e-Commerce	4/19
Vlachos et al.	2011	MSQ	General	3(7)/n.a.
Stiakakis and Petridis	2014	MSQ	General	3(11)/40
Huang et al.	2015	MSQ	General	4-5/15
Kuo et al.	2016	MASQ	m-Commerce	7/35

467f.).

Moreover, considering the results of the SLR, a hierarchical approach is sometimes used to measure ESQ and MSQ (tab. 4). The concept of a hierarchical SQ scale is first introduced by Dabholkar et al. (1996) in the context of retail stores and underpinned with empirical evidence by (Brady and Cronin, 2001). Especially the more recent papers on MSQ incorporate this hierarchical approach to assess the different connotations of SQ (e.g., Stiakakis and Petridis, 2014; Vlachos et al., 2011).

4.1.2. Development

A multi-dimensional and hierarchical scale is often used for measuring ESQ and MSQ as mentioned in the previous subsection. Hence, the scale for measuring MASQ will also apply multiple dimensions to address the issue of MASQ measurement from various perspectives in a comprehensive manner and will use two hierarchy levels for structuring the dimensions. The overall MASQ of a mobile shopping companion app is thus measured using more detailed sub-dimensions and related items for measuring concrete aspects of the SQ.

For determining the primary dimensions of the MASQ scale, the hierarchical scales identified in the SLR are an-

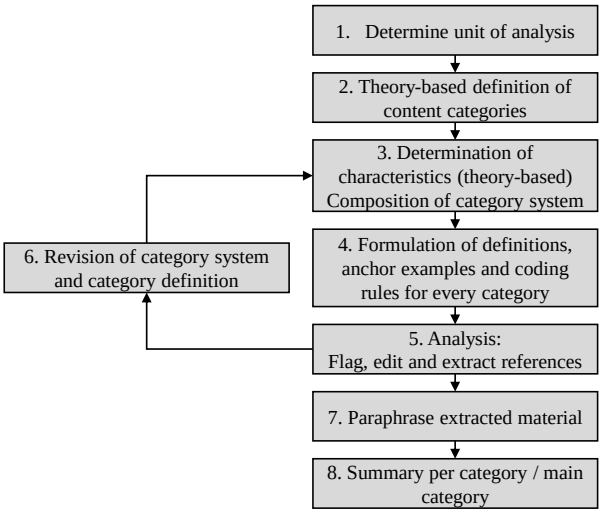


Figure 5: Procedure of the Qualitative Content Analysis (Mayring, 2015, p. 98)

alyzed. Regarding recent hierarchical MSQ measurement scales, it seems that there is at least a consensus concerning the primary dimensions (Fullerton, 2005; Lu et al., 2009; Stiakakis and Petridis, 2014) As primary dimensions interaction quality, environment quality and outcome quality are often used in the context of m- services. These three primary dimensions also form the first-order dimensions for the developed MASQ scale (fig. 6).

To identify relevant secondary dimensions for the MASQ scale, the dimensions of the scales identified during the SLR are compared on a conceptual basis using a conceptual matrix as proposed by (Webster and Watson, 2002). Several authors found that there are some SQ dimensions that are commonly chosen (e.g., Francis, 2009; Hu, 2010). For comparing the dimension, the MSQ dimensions introduced by Stiakakis and Petridis (2014) are used as reference dimensions (tab. 29). These dimensions are chosen as reference because the scale is empirically tested, is developed for measuring MSQ, builds upon prior research on MSQ (e.g., Lu et al., 2009) and provides a comprehensive overview on the SQ of general m-services incorporating all e-Commerce phases. The conceptual comparison of the dimension is not only executed on the denotation of the terms but also considering the specific definitions as presented in literature[6] and related items. Using the conceptual matrix, eight candidate dimensions are identified for measuring the MASQ. These candidate dimensions are commonly used dimensions for assessing the ESQ and MSQ. The percentages in brackets indicate the frequency of occurrence in all SQ scales identified during the SLR (app. C). These are the the secondary dimensions problem solving (84.8%), information (75.8%), security/privacy (78.8%) and customization/personalization (42.4%) subsumed in the primary dimensions interaction quality, equipment (48.5%) and design (78.8%) for the environment quality dimension and reliability (75.8%) and valence (39.4%) for the outcome quality dimension.

The problem solving dimension refers to the quality of the customer support the provider offers in case of any issues. Information resembles to the quality and actuality of the information provided through the mobile app. The dimension security/privacy deals with the protection of the customer's data and quality of security features against vulnerabilities. Customization/personalization describes the ability of a mobile app to be modified according to a user's needs. The equipment dimension is concerned with the performance of the mobile app in context of the mobile device and mobile network. The aesthetics and ease of use are content of the design dimension. Reliability initially describes the reliability of the service completion compared to the agreement with the provider. Finally, the valence dimension measures the customer's overall satisfaction or dissatisfaction with the provided m-service. These dimensions are introduced in more detail following the primary dimension they are related to in section 4.2, 4.3 and 4.4. The remaining secondary dimensions expertise (21.2%), content (9.1%) and tangibles (12.1%) are only rarely used and seem not to be appropriate for measuring the MASQ (tab. 29).

More specifically, the results of the literature synthesis in the conceptual matrix suggest applying modified denotations for the dimensions problem solving and equipment that are used more frequently. Thus, the secondary dimension problem solving is renamed to responsiveness and the secondary dimensions equipment is denoted as performance. The complete conceptual matrix is presented in appendix C.

As the content structuring approach by Mayring (2015) allows to refine the categories established prior to the QCA by using the literature, the candidate dimensions are tested with the actual online customer reviews in a small test sample (sec. 3.2). Analyzing the test sample reveals that the candidate dimension customization/personalization is not of in-

[6] An overview of the dimensions and their definitions of all SQ scales from the SLR can be found in the digital appendix in alphabetic order.

terest for the users of the mobile shopping companion apps. Therefore, this candidate dimension is discarded. Furthermore, the reliability dimensions is further specified, in line with the previous two category refinements and the scope of this thesis. Most of the authors define this dimension as delivery reliability or fulfillment (e.g., Collier and Bienstock, 2006; Francis and Lesley, 2002; Holloway and Beatty, 2008; Huang et al., 2015; Kuo et al., 2016). However, the actual delivery is not part of the MASQ but of the logistics quality that is not included in the MASQ scale. The online customer reviews also tend to highlight technical malfunctions and service dropouts with reliability. Hence, this dimension it further denoted as technical reliability. Moreover, analyzing the test sample, an additional candidate category is inductively found. This category is denoted features and mirrors the influence of a diverse set of features and services on the SQ of a mobile app. Although this dimension appears not to be measurable assessing the mobile app or m-services provided, it is frequently mentioned by customers in the online customer reviews. Thus, it might also be a determinant of MASQ to the regard that a certain feature set is required for achieving a high MASQ. Therefore, the set of features requested by the customers compared with the suggestions from literature is not integrated in the MASQ measurement scale but depicted separately in section 5.5 (e.g., Rudolph et al., 2015).

As even the secondary dimensions include a huge set of quality aspects that can only be hardly measured directly, each secondary dimension is further specified by more detailed items. These items resemble manifestations of the related secondary dimensions that can be measured as determinants of SQ in the context of mobile shopping companion apps. They are either taken from literature or inductively derived from the results of the QCA.

4.1.3. Overview

Having developed the structure of the MASQ measurement scale and shortly introduced the primary and secondary dimensions of the MASQ scale, the final scale for measuring the MASQ is presented in figure 6 and table 5.

The MASQ as overall quality construct is situated on top of the hierarchy at the right (Stiakakis et al., 2013, p. 1). Then, the primary dimensions form the second-level in the middle and the secondary dimensions are located at the bottom of the MASQ hierarchy at the left (fig. 6). Beneath the level of the secondary dimensions, an additional item level is situated that is not displayed here for reasons of clarity. Though, the 22 items that further describe the secondary dimensions are introduced in table 5. This MASQ scale is intended to measure the SQ of the whole m-service process from the information to the after-sales phase as described by Bauer et al. (2006).

4.2. Interaction Quality Dimension

Interaction quality involves the quality dimensions assessing the interaction between the customer and the m-service provider during the service delivery (Stiakakis and Petridis, 2014, p. 942). It resembles the functional dimension introduced by Grönroos (1984) as presented in subsection 2.2.1. It has a significant influence on the overall perceived MSQ, according to Lu et al. (2009). The primary dimension interaction quality is accompanied by the secondary dimensions Responsiveness, Information and Security and Privacy.

4.2.1. Responsiveness

Responsiveness is concerned with the retailer's ability and willingness to help customers by promptly answering their inquiries and solving their problems when interacting with the mobile shopping companion app (Bauer et al., 2006; Ladhari, 2010). Bauer et al. (2006) also consider the "availability of alternative communication channels" (p. 871) as part of the responsiveness dimension. Indeed, alternative channels for communicating with the customer support personnel are necessary if the mobile shopping companion app does not start. Politeness of the customer service personnel is transferred from the traditional service environment to e-services (Parasuraman et al., 1985, 2005). Moreover, several authors also include the retailer's return and exchange policies into these dimensions (e.g., Bauer et al., 2006; Holloway and Beatty, 2008; Parasuraman et al., 2005). However, this is considered to be out of scope of the MASQ scale since these terms and policies are more general in nature and not adapted for the use with a mobile app. Customer service here only refers to issues related with the particular usage of the mobile shopping companion app and not with any logistics related problems such as false or damaged deliveries. Customer service can be required throughout all e-Commerce stages as requests to the retailer can occur at any time in the purchasing process (Holloway and Beatty, 2008, p. 350). Customer service and responsiveness influences the overall perceived SQ but is not useful for predicting the overall SQ because it is not necessarily needed in all mobile purchases (Lee and Lin, 2005; Wolfinbarger and Gilly, 2003). The responsiveness dimension is defined for the present MASQ scale as the retailer's ability to promptly and politely solve a customer's issues related with the mobile app. A high MASQ is achieved if the responsiveness perceived by the customer meets or even exceeds the expected responsiveness. This dimension involves the items customer service availability, problem solving ability, kindness of employees and guidance within the app (tab. 6).

RES1: Customer service availability

Firstly, customer service quality is measured by the item customer service availability. As mobile shopping companion apps allow access to the retailer's m-services without temporal and spatial constraints, customer service also has to be reachable at any time a customer has an issue with the mobile shopping companion app. This instant help can take the form of a communication with an employee via phone or electronic media, e-mail or a Frequently Asked Questions (FAQ) page. Although self-service sites and FAQ pages are available at any time, customers prefer personal advice and communication with a retailer's employee (Santos, 2003, p.

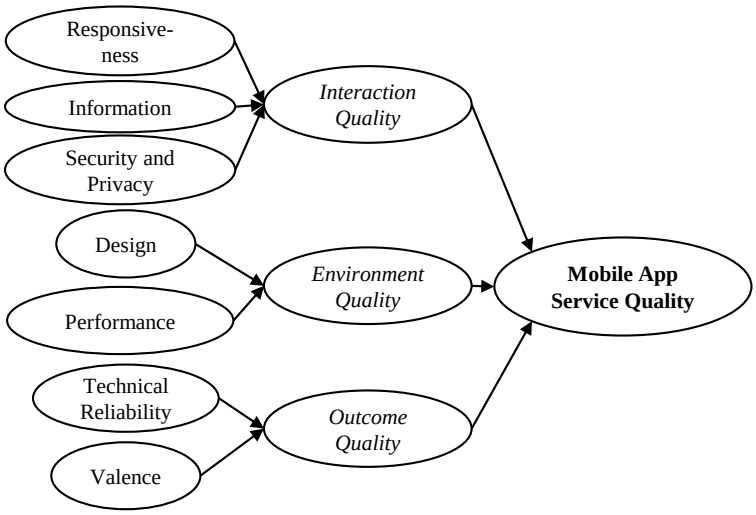


Figure 6: Mobile App Service Quality Measurement Scale

242). The availability of the service personnel can be an issue (Bauer et al., 2006, p. 871).

"I have rung Customer Services on a number of occasions only to be fobbed off if I got one on the line."[7]

RES2: Problem solving ability

Secondly, the retailer's problem solving ability covers prompt and correct responses to customer inquiries. These responses should solve the customer's problems. Advice should be given in a precise, polite and understandable manner giving the customer the feeling that someone is caring for his or her problems.

"Went into store who told me to ring customer services who told me to go on website!!"

RES3: Politeness and kindness of personnel

Thirdly, customer service quality is influenced by the politeness and kindness of the personnel directly interacting with the customer. Analogous to the traditional service environment where Parasuraman et al. (1985) introduced the term "courtesy" (p. 47), customers still expect polite personnel when having issues with their mobile shopping companion app. "I always find Customer Service advisers, polite and caring."

RES4: Guidance and instructions for app usage

Finally, the mobile shopping companion app should incorporate guidance and instructions for using the mobile app. Among other things the actual shopping procedure has to be accompanied with instruction if something remains unclear for the user and features have to be explained for a proper usage. The in-app guidance and instructions for mobile app usage item is introduced to the MASQ measurement scale because customers unfamiliar with the handling of mobile apps begin to make their shopping mobile. Especially, elderly people are introduced to mobile devices by their relatives or friends and try to do their regular shopping with the mobile shopping companion apps. Hence, this item is added to represent these customers. The mobile app should be usable by a new user without considering additional sources of information. This information should be contained within the mobile app.

"Can't find instructions on how to use, and there are features that are not all that intuitive."

4.2.2. Information

The information dimension refers to the quality of information provided through the m- services and the benefit that can be derived from it (Ladhari, 2010, p. 473). The information quality dimension impacts the overall perceived SQ (Collier and Bienstock, 2006; Santos, 2003). It can even affect the customer satisfaction and the customer's reuse and repurchase intentions (Lynch and Ariely, 2000). Information quality is even more important in pure m-services (e.g. online purchase) as they cannot by accompanied by information available in a traditional store (Yang et al., 2005). As mobile shopping companion apps are also designed to enhance the shopping experience in brick-and-mortar stores, they have to sensibly extend the information available offline. The information quality dimension is sometimes integrated within website design or content dimension in existing ESQ and MSQ scales identified during the SLR (e.g., Holloway and Beatty, 2008; Santos, 2003).

For a high SQ, customers demand "up-to-date, detailed, sufficient, and accurate informa- tion" (Stiakakis and Petridis, 2014, p. 943). Yang et al. (2005) differentiate between usefulness and adequacy of information for their information quality dimension (p. 578f.). Information provided through m-services has to be up-to-date at any time. Customers should not feel misled by out-dated information (Santos, 2003, p. 240). Furthermore, the information provided

[7]The online customer reviews used as anchor examples and evidence for implementation guidelines and service design requirements are cited verbatim and are neither corrected in grammar or orthography.

Table 5: Mobile App Service Quality Measurement Scale

Dimension	Description	Sources
Interaction quality	Reflects all the quality characteristics of a customer's interaction with the m-service provider	Lu et al. (2009), Stiakakis and Petridis (2014)
Responsiveness	The retailer's ability to promptly and politely solve a customer's issues related with the mobile app RES1: Customer service availability RES2: Problem solving ability RES3: Politeness and kindness of personnel RES4: Guidance and instructions for app usage	Bauer et al. (2006), Blut et al. (2015), Holloway and Beatty (2008), Parasuraman et al. (2005), Wolfinbarger and Gilly (2003)
Information	The provision of accurate and precise information by the retailer INF1: Information adequacy INF2: Information usefulness INF3: Information correctness	Collier and Bienstock (2006), Fassnacht and Koese (2006), Holloway and Beatty (2008), Loiacono et al. (2007), Santos (2003), Stiakakis and Petridis (2014), Yang et al. (2005)
Security and Privacy	The protection of system and network resources from any external or internal attack and the protection of the customers' personal data SEC1: Information security SEC2: Data protection SEC3: Data collection	Blut et al. (2015), Holloway and Beatty (2008), Parasuraman et al. (2005), Santos (2003), Stiakakis and Petridis (2014), Wolfinbarger and Gilly (2003)
Environment quality	Reflects the context in which mobile apps are delivered, and quality characteristics of the equipment that affect the mobile apps delivery	Lu et al. (2009), Stiakakis and Petridis (2014)
Design	The aesthetics features and layout of the user interface design DES1: Visual aesthetics and clarity of layout DES2: Quality of multimedia content DES3: Ease of use and ease of navigation DES4: Search function and filters	Collier and Bienstock (2006), Fassnacht and Koese (2006), Holloway and Beatty (2008), Parasuraman et al. (2005), Santos (2003), Stiakakis and Petridis (2014), Wolfinbarger and Gilly (2003)
Performance	The performance of the mobile app and its resource requirements PERF1: Processing speed PERF2: Device storage usage and mobile network usage PERF3: Device and connection quality	Collier and Bienstock (2006), Fassnacht and Koese (2006), Huang et al. (2015), Madu and Madu (2002), Parasuraman et al. (2005), Stiakakis and Petridis (2014)
Outcome quality	Reflects the technical quality of the service delivery and the customer's satisfaction with the mobile service	Lu et al. (2009), Stiakakis and Petridis (2014)
Technical reliability	The accurate and consistent operation of the mobile app and the services provided REL1: Mobile app and feature reliability REL2: Availability of m-services REL3: Continuous operation after updating	Fassnacht and Koese (2006), Huang et al. (2015), Li and Leng (2008), Santos (2003), Stiakakis and Petridis (2014), Swaid and Wigand (2009), Wolfinbarger and Gilly (2003)
Valence	The final impression of the customer upon the completion of the service delivery VAL1: Overall satisfaction with the m-services VAL2: Satisfaction with the scope of services	Bauer et al. (2006), Fassnacht and Koese (2006), Loiacono et al. (2007), Stiakakis and Petridis (2014), Tan and Chou (2008), Vlachos et al. (2011)

Table 6: Responsiveness Dimension Summary

Definition	The retailer's ability to promptly and politely solve a customer's issues related with the mobile app
Items	RES1: Customer service availability RES2: Problem solving ability RES3: Politeness and kindness of personnel RES4: Guidance and instructions for app usage
Anchor examples	"I have rung Customer Services on a number of occasions only to be fobbed off if I got one on the line." "Went into store who told me to ring customer services who told me to go on website!!" "I always find Customer Service advisers, polite and caring." "Can't find instructions on how to use, and there are features that are not all that intuitive."
References	Bauer et al. (2006), Blut et al. (2015), Holloway and Beatty (2008), Parasuraman et al. (2005), Wolfinbarger and Gilly (2003)

should be tailored to the customer segment usually purchasing at the retailer to be understandable and relevant (Collier and Bienstock, 2006; Loiacono et al., 2007). The amount of information and its presentation have to be adequate because excessive or substandard information are both perceived negatively by customers (Santos, 2003, p. 240). These characteristics of information quality need to be fulfilled throughout the whole period of interaction with the mobile shopping companion app and need to apply for all information provided such as product, system and retailer information (Fassnacht and Koese, 2006, p. 26).

For the MASQ scale, information quality is defined as the provision of adequate and useful information by the retailer. With regards to the information quality dimension, a high MASQ requires the customers' perceptions of the information quality meet or exceed their expectations. Information quality is further specified by the items information adequacy, information usefulness and information correctness (tab. 7).

INF1: Information adequacy

The information adequacy item is concerned with the "completeness of information" (Yang et al., 2005, p. 579). In case of m-Commerce, complete information involves product descriptions including price transparency and payment information to allow customers to understand the product and make informed purchase decisions. This also involves system information such as the "full disclosure of policies, procedures" (Collier and Bienstock, 2006, p. 264), additional charges and update descriptions including changes, features and bug fixes. Additionally, general information about the provided m-services and the retailer itself should be given (Yang et al., 2005, p. 579).

"Items lack critical input such as dimensions, item #'s, descriptions, etc."

INF2: Information usefulness

Information usefulness can be described by relevancy, clearness, accuracy, consistency and currency. More specifically, information currency involves the timeliness of information and regular updates to incorporate new and adapted information. Information accuracy relates to freedom of errors and information relevancy describes the relevance of information to serve the purpose and task of the customer (Yang et al., 2005; Loiacono et al., 2007).

"Clear and useful product descriptions"

INF3: Information correctness

Information correctness deals with providing and displaying the right information in the mobile app. Among other things, this refers to product descriptions or locations and prices. The correctness of in-store and online stock is essential for the usefulness of the stock availability check. Besides the correctness, the currency is also important as the stock information should ideally be real-time (Holloway and Beatty, 2008, p. 361). This enables customers to see how many items are on stock and if the product is available at their preferred store. The correct and timely stock information is also important for mobile purchases to ensure an on-time delivery. The focus of the information correctness item is placed on the correctness and currency of in-store and online stock information as many customers complain about incorrect stock information in the online customer reviews as presented in the QCA in section 5.1.1.

"I downloaded this app to see if an item was in stock at a store. It says it was. I drove all the way there. It was not there. What good is an app that says something is in stock if it is inaccurate?"

4.2.3. Security and Privacy

Security and privacy is a concern of 26 of the 33 scales identified during the SLR (app. C). Although security and privacy issues are already a concern in e-service environments, Kuo et al. (2016) found that this issue is even of higher importance in mobile shopping apps with less control over personal and payment information (p. 5). Security and privacy haven "a very strong impact on m-service quality" (Stiakakis and Petridis, 2014, p. 943). Issues related to security and

Table 7: Information Dimension Summary

Definition	The provision of adequate and useful information by the retailer.
Items	INF1: Information adequacy INF2: Information usefulness INF3: Information correctness
Anchor examples	"Items lack critical input such as dimensions, item #'s, descriptions, etc." "Clear and useful product descriptions" "What good is an app that says something is in stock if it is inaccurate?"
References	Collier and Bienstock (2006), Fassnacht and Koese (2006), Holloway and Beatty (2008), Loiacono et al. (2007), Santos (2003), Stiakakis and Petridis (2014), Yang et al. (2005)

privacy are relevant across all phases of commerce from information to after sales (Holloway and Beatty, 2008, p. 350).

While Parasuraman et al. (2005) describe privacy as protection of customer-related and credit card information, Holloway and Beatty (2008) interpret this dimension broader, also including the feeling of safety when interacting with an e-service. Customer-related information are a customer's personal information such as name, address and purchasing information that are linked with the customer. Privacy also includes that the retailer is not allowed to hand over customer-related information to any third-party without the customer's explicit permission. In addition, the retailer has to ensure that the customer's personal data and credit card information are secured against any form of fraud or hacker attacks and that the information is not lost due to any system malfunctions, besides the intentional handover of customer-related information (Holloway and Beatty, 2008, p. 357).

Hence, adequate security features of the mobile app are important for a high MASQ such as encrypted data transmission (Fassnacht and Koese, 2006, p. 27) and secured log-in (Wolfinbarger and Gilly, 2003, p. 193). Customers oftentimes do not use the mobile app if they have doubts regarding the security of their personal and credit card information (Blut, 2016, p. 505). For the MASQ scale security and privacy is defined as "the protection of system and network resources from any external or internal attack and the protection of the customers' personal data"(Stiakakis and Petridis, 2014, p. 943). The security and privacy dimension includes the manifestations information security, data protection, data collection (tab. 8).

SEC1: Information security

Information security is not only the most important SQ factor in e-services but also in m-services (Ladhari, 2010; Kuo et al., 2016). Because of the wireless transmission in m- services there is an additional potential security issue. For mobile shopping companion apps, the retailer as developer and distributor of the mobile app has to ensure that personal and payment data that are stored for convenience at the check-out, are kept confidentially without any intended or unintentional access of any third-party or unauthorized personnel. This security has to be assured for the data transmission via the mobile internet, for the storage within the systems of the retailer and especially for the mobile shopping companion app executed on the smartphone. This involves encryption on the device and secured access to the mobile app using passwords or finger prints to restrict at least the access to the check-out to the intended customer if the smartphone is used by another person. The customer should not be in doubt about his or her personal or payment data (Santos, 2003, p. 242).

"Needs better security features was hacked and almost lost my money to a Russian."

SEC2: Data protection

Data protection in this context means that the customer always has to possess the control over his or her personal and payment data. The retailer or its employees are only allowed to store and use the customers' data after they explicitly granted their permission. Especially, for sharing customer-related data with third-parties such as advertisers or suppliers permission is required. This includes explicit indications when data is shared and stored and the possibility to change and delete personal or payment data.

"Without my knowledge or permission, it stored my credit card information. This is absolutely unacceptable!"

SEC3: Data collection

With an increasing safety awareness of the customers, the developers of smartphone operating systems have integrated features to control the permissions of the mobile apps executed on the smartphone. These permissions govern the data collection potential of the mobile app. Among others, these permissions include access to a customer's media files or social media accounts and device functions such as storage, network usage or camera. A mobile shopping companion app is considered as safe in this regard if it requires as less permissions as possible to obtain its complete functionality. This means that for instance the barcode scanning feature of the mobile shopping companion app only works properly if it has the permission to access the camera. Otherwise the screen will stay black. Hence, a potential security breach may be if the mobile shopping companion app requests permissions that are not necessary for a properly execution. This item is newly added to the context of MASQ as data collection and

Table 8: Security and Privacy Dimension Summary

Definition	The protection of system and network resources from any external or internal attack and the protection of the customers' personal data.
Items	SEC1: Information security SEC2: Data protection SEC3: Data collection
Anchor examples	"Needs better security features was hacked and almost lost my money to a Russian." "Without my knowledge or permission, it stored my credit card information. This is absolutely unacceptable!" "The last version of the app [...] requires so many extra permissions that I am not willing to keep installed"
References	Blut et al. (2015), Holloway and Beatty (2008), Parasuraman et al. (2005), Santos (2003), Stiakakis and Petridis (2014), Wolfinbarger and Gilly (2003)

especially permissions are not such an issue in the ESQ environment. The awareness of data collection has evolved in recent years with the emergence of mobile apps on mobile devices.

"The last version of the app [. . .] requires so many extra permissions that I am not willing to keep installed".

4.3. Environment Quality Dimension

Environment quality describes the quality characteristics of the mobile device used to access the m-service and "the quality of the context in which m-services are delivered" (Stiakakis and Petridis, 2014, p. 942). The environment quality has a huge impact on the quality perception compared to e-services because users can use the provided services without spatial and temporal constraints (Heinonen and Pura, 2008, p. 2). The technical features of the mobile device (e.g. screen size, battery capacity) and the network coverage of mobile cellular networks affect the quality of the delivered m-services (Stiakakis and Petridis, 2014, p. 942). Environment quality is measured with the secondary dimensions Design and Performance.

4.3.1. Design

The design dimension is concerned with the visual aesthetics of the User Interface (UI) of the mobile app and features for a more efficient information retrieval such as the internal search engine and filters. Several authors have shown that the visual design of an e-service or m-service highly impacts the customer's overall SQ perception (e.g., Brady and Cronin, 2001; Lu et al., 2009; Stiakakis and Petridis, 2014; Vlachos and Vrechopoulos, 2008; Wolfinbarger and Gilly, 2002). Besides the mere SQ, the design dimension not only influences the customer's intention for reuse and revisit but also the customer's perception of the retailer (Rosen and Purinton, 2004; Sohn and Tadisina, 2008). Although the design dimension is present in 26 of the 35 ESQ and MSQ scales identified during the SLR, there is no consensus about the number of design related dimensions. Parasuraman et al. (2005) use the term "Efficiency" (p. 7) to include design-related aspects in their E-S-QUAL scale. Efficiency describes an e-service as "simple to use, structured properly" (Parasuraman et al., 2005, p. 7) and requiring less input by the customer. Collier and Bienstock (2006) for instance introduce the two dimensions "Design" (p. 264) and "Ease of use" (p. 264) in the context of the design dimension. Santos (2003) even uses the three design-related secondary dimensions "Ease of use", "Appearance" and "Structure & layout" below the "Incubative dimension" (p. 239) in her ESQ measurement scale.

Generally, design refers to the "visual appearance and audible application" (Collier and Bienstock, 2006, p. 264) of a mobile app including "color, animation, pictures, text, format, and sound" (Collier and Bienstock, 2006, p. 264). The visual appearance also includes the layout and structure of the mobile app. It can be used to simplify the use of the mobile app by having a flat and clean design that guides the customer through the whole mobile app (Santos, 2003, p. 239). Design quality is subjectively perceived by every single customer. Preferences for colors used, layouts and images vary widely among customers (Swaid and Wigand, 2009, p. 16). While a specific color is preferred by a certain group of customers, it may be disliked by another. High design quality is achieved if the customer's perceived design quality meets or exceeds their expected design quality. Additionally, a high quality UI should be to some extent innovative (Collier and Bienstock, 2006, p. 272). The design dimension is defined as the aesthetics features and layout of the UI design for the MASQ measurement scale. This dimension encapsulates several facets of the UI of the mobile shopping companion app. Therefore, it incorporates the items visual aesthetics and clarity of layout, quality of multimedia content, ease of navigation and ease of use and search function and filters (tab. 9).

DES1: Visual aesthetics and clarity of layout

Visual aesthetics and clarity of layout describe the customers' overall perception of the "appearance of the site" (Parasuraman et al., 2005, p. 7). This is the first aspect the customer

Table 9: Design Dimension Summary

Definition	The aesthetics features and layout of the user interface design.
Items	DES1: Visual aesthetics and clarity of layout DES2: Quality of multimedia content DES3: Ease of use and ease of navigation DES4: Search function and filters
Anchor examples	"Good size text and clear with suitably neutral colors which are pleasing on the eye." "Pictures are clear and give good detail of the product." "Difficult navigate back to product listings after viewing details of a product." "Good filters allowing easy viewing of appropriate size and fit of items along with other handy filters."
References	Collier and Bienstock (2006), Fassnacht and Koese (2006), Holloway and Beatty (2008), Parasuraman et al. (2005), Santos (2003), Stiakakis and Petridis (2014), Wolfinbarger and Gilly (2003)

observes when he or she starts the mobile shopping companion app (Santos, 2003, p. 239). Thus, it is important to use colors and a structure that suits the majority of the customers. The mobile app should be visually pleasing for the customers (Loiacono et al., 2007, p. 84). While Santos (2003) underpins the importance of the appropriate size for websites in their ESQ scale, this can be transferred to the MASQ scale. Mobile apps should provide an appropriate resolution for the screen size of the mobile device they are executed on without simply enlarging a lower resolution to the higher one. In this regard, providers need to update their mobile apps if a manufacturer introduces larger displays or a higher resolution. The clarity of layout is defined as "the degree to which the design structure of the user interface helps users to find their way" (Fassnacht and Koese, 2006, p. 26). This includes "the organization and presentation [...] of content and information" (Santos, 2003, p. 240). Most importantly the layout has to be simple and clean to foster an easy navigation through the mobile app.

"Good size text and clear with suitably neutral colors which are pleasing on the eye."

DES2: Quality of multimedia content

The item quality of multimedia content measures the quality of the visual representation of multimedia elements in mobile apps such as font, icons and digital images or videos (Fassnacht and Koese, 2006, p. 26). Fonts and icons are considered of high quality if they are easy to read and understand (Loiacono et al., 2007, p. 83). For digital images, it is required that they have an appropriate resolution and are zoom-able enabling the customer to visually capture details of the product. Besides the quality of the graphical elements, the mobile app has to use an appropriate amount of graphics, text, animations and effects because both, too many and too little of each will negatively affect the customer's perceived SQ. Although this item also seems to be relevant for ESQ and MSQ it is not explicitly mentioned in the analyzed literature. Nevertheless, it is added to the MASQ measurement scale to capture the influence of the quality of multimedia content on the perceived SQ. "Pictures are clear and give good detail of the product."

DES3: Ease of use and ease of navigation

In general, the item ease of use and ease of navigation is related to the UI of a mobile app and efficient information retrieval and purchase process (Santos, 2003, p. 239). Ease of use describes the quality of the UI with regards to the facilitation of access to the m-services (Fassnacht and Koese, 2006, p. 26). A customer should be able to find the information required or to process the purchase as efficient as possible. For websites, the concept of as less clicks as possible has been coined (Collier and Bienstock, 2006, p. 264). Concerning m-services, it should be easy and intuitive for the customer to learn how to operate the mobile app (Loiacono et al., 2007, p. 83). In particular, ease of navigation is considered with efficient navigation through the different tabs, sites and products of the mobile shopping companion app. Ideally, customers should always know at which position of the mobile app they are. Further, they have to be capable of getting to any menu level they want as well as to the previous level or site (Collier and Bienstock, 2006, p. 264). A customer should be able to manoeuver through the mobile app back and forth as he or she desires it (Parasuraman et al., 2005, p. 7).

"Difficult navigate back to product listings after viewing details of a product."

DES4: Search function and filters

The search function and filters facilitate the search for specific products by "product, features, or keyword" (Santos, 2003, p. 239). This especially applies for the internal search engine of the mobile app. It should retrieve the information and products the customer searched for. A good search function is required for a high SQ (Parasuraman et al., 2005, p. 7). Filters enable the customers to narrow down the products or search results by specific product characteristics such as color, size or product category etc. While the quality of the search func-

tion is mentioned as determinant for the overall SQ in extant literature, the quality and scope of filters is added based on the analysis of online customer reviews.

“Good filters allowing easy viewing of appropriate size and fit of items along with other handy filters.”

4.3.2. Performance

The performance dimension deals with the overall responsiveness of the mobile app itself. The mobile app should be perceived as smoothly operating without any glitches or dropouts. Several authors mention that the performance dimension is an important aspect for measuring the ESQ (e.g., Collier and Bienstock, 2006; Parasuraman et al., 2005) and the MSQ (e.g., Huang et al., 2015). Kuo et al. (2016) also revealed that efficiency that also encapsulates performance is relevant for measuring MASQ. To fulfill the requirements of m-services that are accessible without spatial and temporal constraints, the m-services provided by mobile shopping companion apps have to be available 24 hours a day to allow customers to shop online even at night (Fassnacht and Koese, 2006, p. 27). The 24 hours availability includes a consistent level of performance the whole day. Customers always expect the same performance when they are using the mobile app (Madu and Madu, 2002, p. 251). It does not matter to them whether the m-services are requested by many customers at the same time as they always have the impression that they are using an individual instance of the mobile app independently of other customers as expressed in the multiplicity dimension by Nickerson et al. (2009) in their taxonomy of mobile apps. Performance is perceived within "the speed of downloading, search, and navigation" and maneuvering through the mobile app (Santos, 2003, p. 241).

High performance quality is achieved if the customer’s perceived performance of the mo- bile app exceeds its expected performance. The performance dimension is defined as the performance of the mobile app and its resource requirements for the MASQ measurement scale. It is refined by the items processing speed, device storage usage and mobile network usage and device and connection quality (tab. 10).

PERF1: Processing speed

Most importantly, the performance dimension involves the processing speed item. It is concerned with the processing performance of any operation within the mobile shopping companion app involving for instance page loading and transitions, smooth scrolling and quick responses to the customer’s inputs (Collier and Bienstock, 2006, p. 264). Processing speed is also concerned with the "ability to quickly download information" (Madu and Madu, 2002, p. 251) such as visual graphics from the mobile app. Besides the processing speed it is also related to quality of data processing and transfer (Fassnacht and Koese, 2006, p. 27).

“There are a few long pauses when switching the screens.”

PERF2: Device storage usage and mobile network usage

The item device storage usage and mobile network usage measures the amount of disk space the mobile app requires on the mobile device and the network traffic it causes for the mobile cellular network and the customer’s purchased volume of data. Madu and Madu (2002) introduce a "storage capability" (p. 251) dimension for e-services. However, this dimension is more concerned with the capability of the website to store information and the amount of information it is able to store. In the context of mobile apps, the device storage usage is concerned with the disk space usage on the mobile device. As the storage space of a mobile device is finite and cannot be extend for all devices with additional memory cards, mobile apps should have an appropriate size. Appropriate in this case means that the mobile app requires as little storage as required for the provisioning of its m-services to the customer. The memory usage is a major concern for not downloading a mobile app provided by a retailer (Lipsman, 2017, p. 40). Mobile network data usage measures the traffic the mobile app causes (Knote and Söllner, 2017, p. 129). Although this item is not derived from the SLR, it is mentioned frequently in the QCA and also inductively derived in another paper (Knote and Söllner, 2017, p. 125). Especially, for volume-based and per-megabyte contracts it is important that the mobile app does not cause to much avoidable mobile network traffic besides the traffic required to provide its features and information. Avoidable mobile network traffic can be caused among other things by constantly fading-in advertisements or uncompressed data transfer.

“It runs in the background, and uses a LOT data!!”

PERF3: Device and connection quality

Stiakakis and Petridis (2014) introduced an "Equipment" dimension that is concerned with the performance of the customer’s mobile device itself and the quality of the mobile internet connection (p. 943). However, the device and connection quality dimension is not directly related with the mobile app itself, it can influence the performance and thus the perceived SQ of the mobile app because an error of the mobile device or malfunctions and glitches that are caused by a connection interruption are likely to be projected on the SQ of the mobile app. Several authors have proven that this item is related to the primary dimension environment quality (e.g., Lu et al., 2009; Vlachos and Vrechopoulos, 2008).

"I had this app on my Motorola phone and it scanned the receipts without any problems. When I change to a Samsung the app stopped scanning."

4.4. Outcome Quality Dimension

Outcome quality reflects the technical quality of the service delivery and the customer’s satisfaction with the mobile service. It is similar to the technical quality dimension introduced by Grönroos (1984). In contrast to Stiakakis and Petridis (2014) and Lu et al. (2009), it does not involve the quality of the product delivered in the context of m- Commerce and only considers m-services provided through a mobile shopping companion app. As customers are likely to evaluate the quality of the whole service endeavour jointly (Li and Leng, 2008), outcome quality also plays an important role in measuring the overall perceived MASQ. The outcome

Table 10: Performance Dimension Summary

Definition	The performance of the mobile app and its resource requirements.
Items	PERF1: Processing speed PERF2: Device storage usage and mobile network usage PERF3: Device and connection quality
Anchor examples	“There are a few long pauses when switching the screens.” “It runs in the background, and uses a LOT data!!” "When I change to a Samsung the app stopped scanning."
References	Collier and Bienstock (2006), Fassnacht and Koese (2006), Huang et al. (2015), Madu and Madu (2002), Parasuraman et al. (2005), Stiakakis and Petridis (2014)

quality is described using the secondary dimensions Technical Reliability and Valence.

4.4.1. Technical Reliability

In the context of IS, reliability is defined as consistent performance according to predefined specifications under certain boundary conditions (Fischer and Hofer, 2008, p. 698). Reliability is one of the quality properties mentioned in chapter 2.1 and describes the "probability of the system executing without failure for a specific period of time" (Pohl, 2010, p. 19). The promised services should be operated "accurately and consistently" (Santos, 2003, p. 241) in line "with the provider’s promises/guarantees" (Stiakakis and Petridis, 2014, p. 944) and the mobile app should operate properly without any crashes, glitches or malfunctions (Huang et al., 2015, p. 134). In general, the perceived reliability is a major determinant for the overall SQ (e.g., Fassnacht and Koese, 2006; Lu et al., 2009; Santos, 2003; Seth et al., 2007). This dimension focuses on the technical aspects of reliability in the context of mobile apps. It is only concerned with technical aspects that can be directly related to the mobile shopping companion app and m-services directly related to it. Among other things, these are the consistent performance of the overall mobile app, the accuracy of in-app payment and billing and the correct operation of single features such as the barcode scanner (Santos, 2003, p. 241). Hence, neither any logistics-related quality aspects such as delivery time or delivery quality (Lin et al., 2016, p. 392) nor assortment-related aspects such as quality and diversity of products (Wolfinbarger and Gilly, 2003, p. 189) are in scope of this dimension.

Although the composition of this dimension is unique compared to the extant literature, most of the identified papers subsume aspects of technical reliability under various dimensions. Parasuraman et al. (2005) introduce the technical reliability as correct service operation within the "System Availability" (p. 18) dimension. Lu et al. (2009) partially integrate aspects of this dimension in their "punctuality" (p. 233) dimension that emphasizes the timely delivery of a service including the logistics quality. In contrast to the narrow description of reliability as timely delivery, Stiakakis and Petridis (2014) use the broader conceptualization of reliability that incorporates the punctuality aspect raised by Lu et al. (2009). Li and Leng (2008) integrate the technical reliability dimension within their "assurance" (p. 4) dimension that combines delivery reliability, responsiveness and update frequency. Besides the accurate and consistent operation of the mobile app, the technical reliability dimension also includes frequent updates for the mobile apps (Santos, 2003, p. 241). As content updates (e.g. new products or product reviews) are included within the information quality dimension, updates in the context of the technical reliability dimension incorporate technical updates of the mobile app such as adding, improving or removing features, fixing technical problems or adapting security mechanisms. In addition, the mobile shopping companion app has to correctly handle the purchases made by a customer and transfer the purchase data to the retailer using certain m-services. The transaction should be free of errors from a technical point of view (Wolfinbarger and Gilly, 2003, p. 188). Moreover, automatically generated confirmation e-mails (e.g. for confirming a purchase) are considered part of the technical reliability dimension (Swaid and Wigand, 2009, p. 16).

However some authors claim that the reliability quality can only be measured after the completion of the m-service including the delivery and payment (e.g., Fassnacht and Koese, 2006; Stiakakis and Petridis, 2014), the technical aspects of reliability does not include the delivery and thus can be perceived at any time of the service delivery. For instance, a malfunction or crash of a certain feature can be perceived by the customer immediately within the operation because a completion of the service is often impossible after a crash. A high technical reliability is achieved if the perceived technical reliability of the overall mobile app meets or exceeds the customer’s expectations. Nevertheless, the QCA shows that aspects of technical reliability are mentioned if a provided feature or service does not work properly. In the context of the MASQ measurement scale, the technical reliability dimension is defined as the accurate and consistent operation of the mobile app and the services provided. This dimension is subdivided into the items mobile app and feature reliability, availability of m-services and continuous operation after updating (tab. 11).

Table 11: Technical Reliability Dimension Summary

Definition	The accurate and consistent operation of the mobile app and the services provided.
Items	REL1: Mobile app and feature reliability REL2: Availability of m-services REL3: Continuous operation after updating
Anchor examples	“Have to manually add receipt information after the last update. The camera comes on but doesn’t focus or scan the Barcode or QR code.” "Used to be able to access my sparks but now I just get an error message saying sparks are not available." “Where is search – latest update seems to have removed it?”
References	Fassnacht and Koese (2006), Huang et al. (2015), Li and Leng (2008), Santos (2003), Stiakakis and Petridis (2014), Swaid and Wigand (2009), Wolfinbarger and Gilly (2003)

REL1: Mobile app and feature reliability
The mobile app and feature reliability item is used to measure the promised performance and proper operation of the overall mobile app and single features. Often, the mobile app reliability is related to the proper start and continuous operation of the mobile app without any glitches. In extreme cases, glitches and malfunctions can lead to a crash of the mobile app requiring a restart involving the risk of losing previous information such as selected products or already entered data. Rarely, the QCA reveals that the mobile shopping companion app does not even start because of any technical inconveniences so it is completely useless for the customer. Besides, also single features cannot operate properly due to technical reasons such as the online check out or product search. The malfunction of a single feature can also cause the crash of the whole mobile app.

“Have to manually add receipt information after the last update. The camera comes on but doesn’t focus or scan the Barcode or QR code.”

REL2: Availability of m-services
The availability of m-services is an item of the technical reliability dimension because it is difficult for a customer to differentiate if a malfunction is caused by a glitch in the mobile app or an unavailability of a requested m-service. Availability is another quality property of a system and describes "the percentage of time during which the system is actually available" (Pohl, 2010, p. 19). For the proper operation of the mobile app the m-services have to be available and the mobile app itself has to be to the greatest possible extent free of errors. The m-services provided through the mobile app have to be available at any time a customer wants to access and use them. As m-services can be accessed without temporal and spatial constraints, they have to be available at any time. An unavailability of an m-service negatively influences the perceived technical reliability of the mobile app. "Used to be able to access my sparks but now I just get an error message saying sparks are not available."

REL3: Continuous operation after updating
As frequent updates are a concern of the technical reliability dimension, the continuous operation after updating has to be ensured. The updates should be applied with the objective of improving the customer’s perceived SQ and providing additional or required m-services. For each update, there should be a detailed update description or release notes to inform the customer about any intended changes. If features are removed, there should be a description giving evidence for the feature deletion. Above all, the accurate and consistent operation has to be taken for granted after the update. This includes the proper operation of all features and m-services as beforehand the update including the storage and reloading of the customer information.

“Where is search – latest update seems to have removed it?”

4.4.2. Valence

The valence dimension describes the customer’s feeling after having experienced the m- services through the mobile shopping companion app. This feeling can be anything between "good or bad" (Stiakakis and Petridis, 2014, p. 944) after the completion of the service endeavor. A good feeling can be created for the customer by enabling him or her to efficiently use the mobile app in the shortest possible amount of time. Most importantly, the customer has to gain positive experiences when using the mobile app. These positive feelings may result in an overall positive experience of the mobile app and a positive judgment regarding the SQ of the mobile app and the retailer itself (Stiakakis and Petridis, 2014, p. 947). The valence dimension is concerned with emotional and hedonic aspects of SQ mostly influencing the overall customer satisfaction (Stiakakis and Petridis, 2014, p. 944). A customer that is enjoying the website or mobile app provided by the retailer is likely to spend more time and money on it than expected (Fassnacht and Koese, 2006, p. 27). Although, the "enjoyment dimension" (Bauer et al., 2006, p. 873) is only rarely reflected in the existing literature, Bauer et al. (2006) found that it has huge impact on customer profitability and re-purchase intentions.

The valence dimension is at least partially addressed in

several identified papers using var- ious denotations (e.g., Bauer et al., 2006; Cai and Jun, 2003; Loiacono et al., 2007; Vlachos et al., 2011). In line with the positive experience that a high SQ mobile app should create, Fassnacht and Koese (2006) introduce the "emotional benefit" (p. 27) dimension. It is defined as "the degree to which using the service arouses positive feelings" (Fassnacht and Koese, 2006, p. 27). Li and Leng (2008) also introduce the emotional benefit dimension and follow the aforementioned definition by Fassnacht and Koese (2006). Another dimension including aspects of valence is the "trustworthiness" (Cai and Jun, 2003, p. 512) dimension. It measures the customer's perceived feelings of the online purchase and its relation to the trust in the retailer. Vlachos et al. (2011) incorporate aspects of the valence dimension within usefulness. An enhancement of performance when using m-services positively influences a customer's feelings (p. 38). Tan and Chou (2008) incorporate perceived playfulness in their "perceived usefulness" (p. 652) dimension. Perceived playfulness also measures the feelings and emotions in the e-Commerce environment.

Every buying decision is determined by a combination of benefit dimensions. Besides utilitarian benefits there are also hedonic and emotional benefits as determining dimensions (Sheth et al., 1991, p. 27). Bauer et al. (2006) found that online shopping can be fun, exciting and entertaining (p. 870). Loiacono et al. (2007) revealed that mobile shopping can make customers feel happy, cheerful and sociable when using mobile shopping apps (p. 84). These feelings influence the actual and future buying decision and even the relation with the retailer (Stiakakis and Petridis, 2014, p. 947). A high valence is achieved if the customers' perceived feelings after the completion of the transaction using a mobile app meet or exceed the customer's expectations before the use of the mobile app. For the MASQ measurement scale, valence is defined as "the final impression of the [customer] upon the completion of the service delivery" (Stiakakis and Petridis, 2014, p. 944). In contrast to the previous secondary dimensions, valence cannot be measured by analyzing the mobile app itself. It can only be measured by asking at least a sample of customers about their satisfaction with the mobile shopping companion app. The customers' opinion can be retrieved for instance in personal interviews and questionnaires or by analyzing online customer reviews. This dimension is expressed by the items overall satisfaction with the m-service and satisfaction with the scope of services (tab. 12).

VAL1: Overall satisfaction with the m-service

The first item measures the overall satisfaction with the m-service after the usage of the mobile shopping companion app and the m-services it provides access to. It also reflects the impact of the valence dimensions on the reuse intention of the mobile app and the repurchase intention at the retailer providing the mobile shopping companion app. In addition, it influences the customer's intention to recommend it to other customers. Customers having positive feelings after using the m-services through the mobile shopping companion app are likely to use the mobile app again and repurchase at this particular retailer. Contrary, dissatisfied customers will not reuse the mobile app and are likely to switch to a competing retailer.

"This method of viewing items and proceeding to the payment section is the easiest possible way of ordering online that I have come across."

VAL2: Satisfaction with the scope of services

Besides the overall satisfaction with the m-services, the next item satisfaction with the scope of services measures the customers' satisfaction with the set of features and m- services provided by the mobile shopping companion app. This item is neither adapted from the ESQ nor the MSQ environment, but created for the MASQ measurement scale based on the QCA that reveals that some online customer reviews deal with the provided scope of m-services by the mobile shopping companion app. The set of features provided by mobile shopping companion apps is already introduced in subsection 2.3.2 and the features additionally requested are subsumed within the features dimension. Although each retailer should decide upon the implemented features and m-services based on its group of customer's demands and objectives of the mobile app, a proposed set of features and m-services for an exemplary mobile shopping companion app of an arbitrary retailer is deduced in section 5.5. Nevertheless, retailers seem to follow two distinct approach for the assembly of their mobile apps, either implementing one mobile app encapsulating all provided m-services or providing smaller sets of m-services in single apps (sec. 5.4.2). "You have so many apps doing different things! Payqwiq, Clubcard and Groceries. Link them for easy use."

5. Design Requirements for Mobile Shopping Companion Apps

The design requirements that will lead to a high MASQ for mobile shopping companion apps are presented in the following chapter. This chapter is structured according to the MASQ measurement scale developed and presented in chapter 4 using the identified primary and secondary dimensions from the SLR and QCA as a framing. To underpin the derivation of the design requirements from the online customer reviews, exemplary extracts from the online customer reviews are quoted within this chapter. Selected design requirements are further illustrated with screenshots taken from the three exemplary mobile shopping companion apps.

5.1. Overview of Analyzed Customer Reviews

Before coming to detail with the design requirements that will lead to high MASQ, the three mobile shopping companion apps chosen as reference for this thesis and the retailers providing them for their customers are introduced. As the design requirements are derived from online customer reviews, some statistics about the sample of online customer reviews exported from the MADPs for the QCA are presented as well.

Table 12: Valence Dimension Summary

Definition	The final impression of the customer upon the completion of the service delivery.
Items	VAL1: Overall satisfaction with the m-services VAL2: Satisfaction with the scope of services
Anchor examples	"This method of viewing items and proceeding to the payment section is the easiest possible way of ordering online that I have come across." "You have so many apps doing different things! Payqwiq, Clubcard and Groceries. Link them for easy use."
References	Bauer et al. (2006), Fassnacht and Koese (2006), Loiacono et al. (2007), Stiakakis and Petridis (2014), Tan and Chou (2008), Vlachos et al. (2011)

5.1.1. Exemplary Retailers and Mobile Shopping Companion Apps

The three reference mobile shopping companion apps are provided by Walmart, Tesco and M&S. They are chosen as exemplary retailers as they address a large proportion of people in the countries of their operation and the majority of the online customer reviews for their mobile apps is available in English. Moreover, these retailers operate traditional brick-and-mortar stores, introduced online shops and most recently launched mobile shopping companion apps for improving their customers in-store and online shopping experience. The mobile shopping companion apps are distributed via MADPs as native apps executed directly on the customer's smartphone accessing the hardware via application programming interfaces and libraries (Charland and Leroux, 2011).

Walmart Stores, Inc.

As an American multinational retailer, Walmart operates a total of 11,695 stores in 28 countries employing more than 2.3 million people worldwide (Walmart, 2017b, p. 16). Walmart was founded in 1962 by Sam Walton in Rogers, United States of America. Today, Walmart's headquarter is still situated in Rogers (Walmart, 2017a). In the fiscal year 2017 Walmart generated a total revenue of USD 485.9 billion and a net income of USD 14.3 billion (Walmart, 2017b, p. 36). Walmart's mobile shopping companion app is called "Walmart: In-Store & Online Shopping. Easy Reorders". It was relaunched in the current form as an upgraded mobile shopping companion app in 2011 (Perez, 2011).

The Walmart app encapsulates all m-Commerce related services within one mobile app. It enables the customer to access the online shop without temporal and spatial constraints, allows to find the nearest Walmart store including stock availability, has an integrated administration for loyalty cards and mobile payment and allows to access the Walmart savings catcher (app. B). The start screen of the Walmart app is shown in figure 7.

The savings catcher is a unique selling proposition of Walmart's mobile shopping companion app compared to the other two mobile apps. The savings catcher compares the prices of products on the customer's receipt with the prices in other Walmart stores and other retailers in the area and offers a cash back if a cheaper price is found in the form of an electronic gift card in the amount of the price difference. The cash back can also be accumulated within the savings catcher and payed out at a later point in time. To promote for a possible cash back, the customer either has to scan a quick response code on the bottom of the paper-based receipt or has to type in the receipt number manually. The purchase and the receipt should not be older than seven days. Afterwards, it is checked whether a lower price can be found by the mobile shopping companion app (Walmart, 2017c). The mobile app highlights the lower price with a green cash back icon and indicates the amount of cash back and the receipt in an electronic form that causes the cash back (fig. 8).

Tesco PLC

Tesco is a Britain-based retailer operating multinationally in twelve countries with 6,500 stores employing 460,000 people worldwide (Tesco, 2017c, p. 1). It was founded in 1919 by Jack Cohen as a group of market stalls (Tesco, 2017b). Its headquarter is situated in Welwyn Garden City, England (Tesco, 2017a). Tesco closed the last fiscal year with a total revenue[8] of USD 72.4 billion and a net income of minus USD 70 million (Tesco, 2017c, p. 1). In contrast to e.g. Walmart providing an integrated mobile shopping app incorporating all m-Commerce services, Tesco provides several m-Commerce related mobile apps e.g. for finding the shelf number (Tesco Finder), for shopping wine (Wine Finder), for online payment (Tesco PayQwiq), or for managing Tesco's loyalty card (Tesco Clubcard). Besides these mobile apps, Tesco provides a "Groceries" app enabling customers to purchase groceries without temporal and spatial constraints (fig. 9). This particular mobile app is used as reference mobile app from Tesco because it provides the largest set of features compared to the simpler mobile apps mentioned before including the access to Tesco's online shop (Brown, 2010). The Tesco Groceries app allows for searching products, arranging delivery time slots, create shopping lists and syncing orders with the customer's online account (app. B). The Tesco Groceries app was released in 2010 (Brown, 2010).

[8]The figures are converted from GBP to USD for a better comparability.

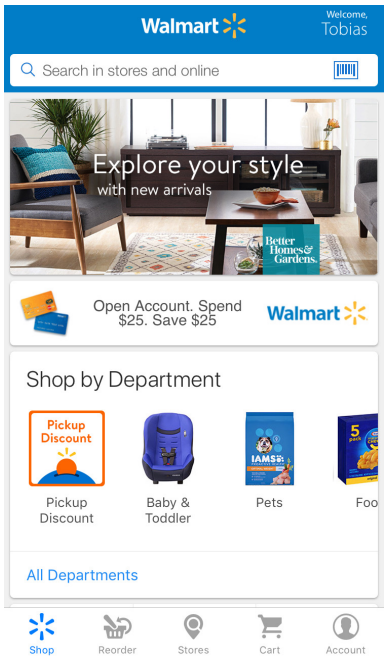


Figure 7: Walmart App Start Screen

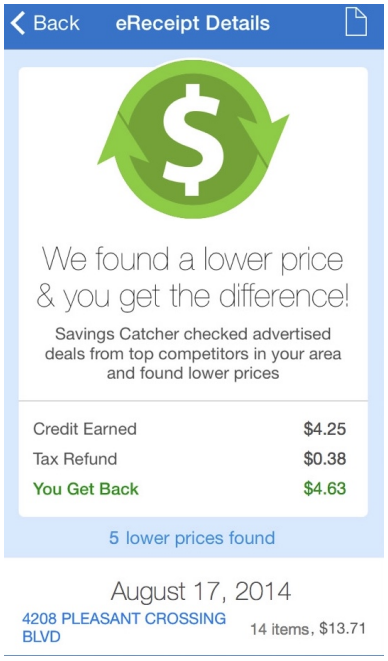


Figure 8: Walmart Savings Catcher (148apps.com, 2017)

Marks and Spencer plc

M&S also is a Britain-based multinational retailer operating in 55 countries and 1,433 stores (M&S, 2017a). M&S employs 85,000 people in total (M&S, 2017b, p. 108). Founded in 1884 by Michael Marks and Thomas Spencer, M&S is nowadays headquartered in the City of Westminster, England (M&S, 2017c). The fiscal year 2017 resulted in a revenue8 of USD 13.7 billion and a net income of USD 152.8 million for M&S (M&S, 2017b, p. 8). The mobile shopping companion app provided by M&S is also called "M&S". M&S states that already 52 percent of all online orders are processed using the mobile app on mobile devices (M&S, 2017b, p. 7). M&S' mobile shopping compan ion app integrates the management of their loyalty program Sparks, notifications for new products and offers and access to the online shop (app. B). The start screen showing the clothing section of the M&S

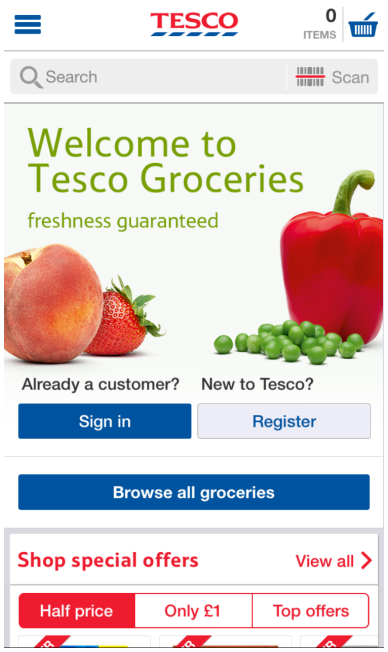


Figure 9: Tesco App Start Screen

app is presented in figure 10. For fostering its multi-channel strategy M&S introduced its mobile shopping companion app in 2012 (Faulkner, 2012).

For further information about the three particular mobile shopping companion apps and other m-Commerce-related mobile apps provided by the three retailers see appendix B.

5.1.2. Online Customer Review Statistics

The MASQ is perceived subjectively by every single customer who uses a mobile app. Thus, the customers' perceptions, experiences and expectations need to be used to develop design requirements for a high MASQ for mobile shopping companion apps. Therefore, online customer reviews are analyzed as a form of eWOM in a QCA to develop these design requirements from the perspective of the customers as potential users. By doing so, it is ensured that the design and feature set of the mobile app fits the customers' desires. A total of 10,099 customer reviews has been exported from the two MADPs App Store and Play Store for the QCA on the 3rd of July 2017 (tab. 13). The online customer reviews evaluate one of the three mobile shopping companion apps by Walmart, Tesco9 or M&S. The extracted online customer reviews cover a time period from September 2014 to the beginning of July 2017. Due to the fact that mobile apps are updated frequently within short time intervals (Nayebi et al., 2016), for the analysis only 8,237 online customer reviews from 2017 are considered (tab. 13).

The sample of online customer reviews has an overall rating of 4.1 stars across all mobile shopping companion apps[9] and MADPs. Following, the ratings of each mobile shopping companion app are presented including the overall rating of the three mobile shopping companion apps in brackets taken from the MADPs (app. B). The mobile shopping companion app provided by M&S is rated highest within the sample 4.5 (4.0) followed by the mobile app by Tesco 4.4 (4.5) and Walmart 3.4 (3.0) for the App Store. In the Play Store the mobile shopping companion app implemented by Walmart is rated with 4.2 (4.5) and the mobile app by M&S with 4.0 (4.3). The average rating of the sample of online customer reviews differs only slightly from the overall ratings in the MADPs. In the initial sample of online customer reviews, each review tends to be very short with an average of 89 characters, similar to Pagano's and Maalej's (2013) findings. As the Play Store allows to publish empty customer reviews, the total length of the online customer reviews varies from zero to 2,705 characters.

In line with PPagano and Maalej (2013), the current analysis reveals that the average customer only posts a single online customer review in the MADP. Contrary, a small number of customers rates a mobile app several times. One customer rated a mobile app nine times over the time period of analysis. Admittedly, this finding has to be treated carefully, as the actual customer can hide himself or herself behind nicknames in the MADP. Hence, the online customer reviews cannot be related to a real customer as they are enabling a single customer to publish reviews under a number of nicknames (Pagano and Maalej, 2013, p. 128).

Following Knote and Söllner (2017) online customer reviews useless for the derivation of design requirements are discarded (sec. 3.2). More specifically, online customer re-

[9]Because of restrictions in the online crawler service, for Tesco only reviews from the App Store could be exported.

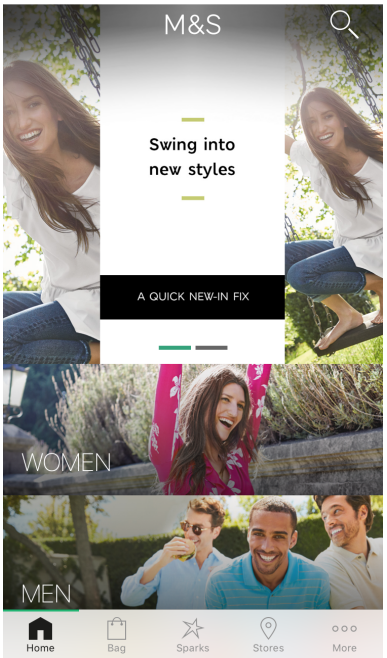


Figure 10: M&S App Start Screen

Table 13: Number of Analyzed Customer Reviews

	Walmart	M&S	Tesco
Number of exports from App Store	1,084	818	662
Number of exports from Playstore	4,964	2,571	n.a.
Total number of reviews		10,099	
Total number of reviews from 2017		8,237	
Total number of relevant reviews		1,448	

views are discarded from the QCA if their topic is not concerned with the SQ of the mobile shopping companion app itself but other retailer related topics. Reviewers also tend to complain or praise the retailer's general assortment ("I would recommend all Marks wine I have yet to find a bad one. They are all equally wonderful."). Sometimes the customers also evaluate the product quality in the online customer reviews ("I like the quality of your clothes"). Or they express their praise for the retailer ("I enjoy shopping at that stores"). Sometimes it is not even obvious if the mobile shopping companion app, the retailer's employees or the appearance of the brick-and-mortar stores is meant by an online customer review ("Always clean, friendly, helpful!"). Excluding these online customer reviews, the manual process results in 1,448 online customer reviews further analyzed using the QCA approach introduced by Mayring (2015).

As most of the online customer reviews cover more than one specific topic, the online customer reviews are coded using the secondary dimensions and items identified in chapter 4 allowing for multiple assignments for a single online customer review and sub-classifying a single online customer review into its components (Pagano and Maalej, 2013, p. 129). This results[10] in a total of 1,307 coded components of online customer review that are distributed as follows among the seven secondary dimensions: Responsiveness (93), Information (103), Security/Privacy (67), Performance (140), Design (452), Technical Reliability (394) and Valence (58) (fig. 11). The QCA also shows that nearly 400 online customer reviews deal with features demanded by customers for a mobile shopping companion app. The resulting feature set for mobile shopping companion apps is the concern of section 5.5.

To allow for a more detailed description of design requirements for mobile shopping companion apps, the components of the online customer reviews are further coded according to the items specifying each secondary dimension. The amount

[10]The raw exports of the online customer reviews from the MADPs and the results of the analysis including the coding for each online customer review can be found in the digital appendix.

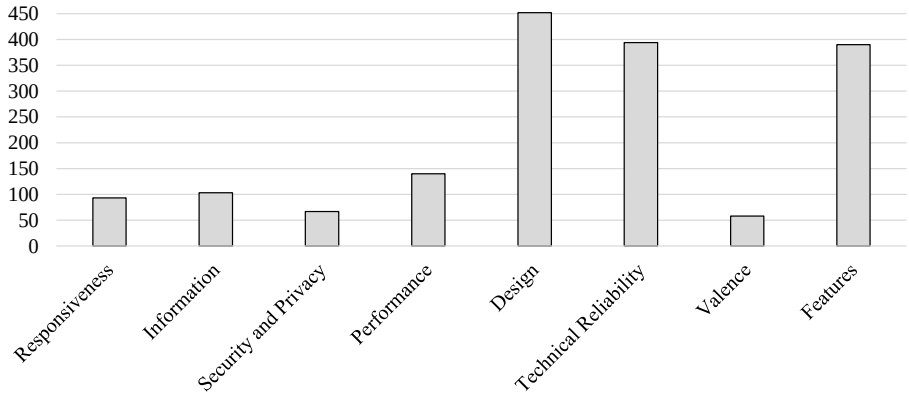


Figure 11: Online Customer Review Statistics - Dimensions

of online customer review components per item varies between five and 298 per item (tab. 14). This fact already highlights that the customers emphasize some of the dimensions and items with higher importance than others. Ease of use and ease of navigation as determinant of the design dimension seems to be more important to the customer than for instance the politeness and kindness of personnel in the customer service specifying the responsiveness dimension (tab. 14).

Although the scope of this thesis lies on the development of a MASQ measurement scale and design requirements for the improvement of the SQ of mobile shopping companion apps, the QCA of the online customer reviews reveals some further characteristics that are also worth mentioning briefly. Firstly, the online customer reviews taken from the App Store tend to be longer and more detailed in total than the reviews in the Play Store that often consist only of a single or a few words. In the extracted sample, the average online customer review from the App Store has a length of 143 characters while the average review from the Play Store consists of 70 characters and also allows for empty reviews (digital appendix). Some users explicitly mention that they like the convenience of the mobile shopping companion app over website that can be accessed using a mobile device without installing a dedicated particular mobile app. Exemplary, a user stated that using the mobile app is "better than just looking at the mobile version of the website!".

5.1.3. Implementation Guidelines and Design Requirements Overview

From the analyzed online customer reviews, 22 implementation guidelines and 14 service design requirements are derived. They are depicted in table 15. While the implementation guidelines directly refer to the mobile shopping companion app suggesting possible changes in the actual implementation of the mobile app, the service design requirements affect the design of the m-services provided by the retailer. These 36 suggestions for adapting the mobile shopping companion app and related m-services should help retailers and their employees to improve the MASQ of their provided mobile apps. The presented suggestions resemble possible areas for improving the MASQ.

5.2. Interaction Design Requirements

Interaction design requirements include all the design requirements derived during the QCA that are related to the interaction quality primary dimensions. While the service design requirements are generally formulated and related to the m-services, the implementation guidelines make concrete suggestions for improving the mobile shopping companion apps. The subsections are labeled analogous to the secondary dimensions of the MASQ measurement scale as presented in section 4.2 as Responsiveness, Information and Security and Privacy.

5.2.1. Responsiveness

The responsiveness design requirements deal with the customers' demands for a customer support providing a high SQ involving the retailer's employees and the mobile shopping companion app (tab. 16).

R01: Continuous availability

Most importantly, the customer service should be available continuously, at any time a customer needs to contact the customer service no matter the reason for the need for a contact. Moreover, the customer service availability and its responsiveness should not be limited due to a high volume of customers contacting the customer service simultaneously. "When you call about the issue a message states that there is a high call volume and tells you to call back and hangs up on you. Seems Walmart doesn't really care about what is going on and why customers keep receiving these emails."

G01: Provision of contact details

The establishment of the contact should be as easy as possible for the customer and free of charge. Contact details for the customer support should not only be made available in the mobile shopping companion app but should also be presented on the retailer's website, in the brick-and-mortar stores and on the general hotline. The mobile shopping companion app should include contact details for specific retailer

Table 14: Online Customer Review Statistics - Items

Dimension	Items			
Responsiveness	RES1: 27	RES2: 52	RES3: 5	RES4: 9
Information	INF1: 29	INF2: 28	INF3: 46	
Security and Privacy	SEC1: 35	SEC2: 16	SEC3: 16	
Design	DES1: 38	DES2: 20	DES3: 298	DES4: 96
Performance	PERF1: 119	PERF2: 10	PERF3: 11	
Technical Reliability	REL1: 312	REL2: 19	REL3: 63	
Valence	VAL1: 45	VAL2: 13		
Features	FET1: 145	FET2: 245		

and customer related topics such as complain handling, return or warranty issues. figure 12 shows the page of the M&S app for gathering the contact information of the retailer or directly get into contact with M&S' employees. "No one to call no one to email."

R02: Multiple contact channels

A high SQ mobile shopping companion app should provide multiple channels for the establishment of a contact to the customer service besides the mobile app. The interaction with the customer service personnel should be possible using for instance phone or e-mail and by consulting a retailer's employee directly in the store. This requirement is already introduced for e-services (Bauer et al., 2006, p. 871).

"Plus if it is weekend no help even reaching out to store manager, they would have to call same number"

G02: Integration of customer service chat

As multiple contact channels are required by the customers, the mobile shopping companion app itself should at least offer a dedicated channel to contact the customer service personnel. The Walmart and M&S apps integrate a chat to enable immediate customer support. Another possible communication channel would be the integration of a more sophisticated video chat allowing for virtual face-to-face communication similar to traditional service environments. The landing page for starting the customer service chat including the possibility to choose a specific topic for the M&S app is depicted in figure 12.

"Had to start a CS chat twice in 3 days because it wouldn't allow me log in or change my password without the chat rep sending me a code."

R03: Satisfaction with problem solving ability

The customer service personnel should be able to assist the retailer's customers with any inquiries or problems. The inquiries and problems should be solved to the customer's satisfaction and the established customer contact should only be terminated if the customer is satisfied with the offered solution. An unsatisfactory solution in the context of m-Commerce is proposing to use the website as an alternative, if the mobile shopping companion app is not working because the mobile app becomes obsolete in case the customer has to use the website anyhow.

"Nobody in the store or at customer service when I call can help."

As the mobile shopping companion apps are deployed by multi-channel retailers operating brick-and-mortar stores, the first contact person in case of issues with the mobile app is not the development or customer service personnel in many cases but the retailer's employees available at the local stores such as cashiers or shopping assistants. Thus, they have to be trained in using the mobile shopping companion apps and providing at least first-level support to customers with complaints and issues.

"The employees are clueless on how to even help you."

Moreover, the customer service should not forward the customer continuously to other contact points if it is avoidable. If a problem cannot be solved by the first contact person, the customer's issue should be forwarded to the technical support or a contact to the technical support should be established by distributing the contact information of the technical support. The contact information of the customer service has to be unambiguous and establish a single point of contact for the customer so that he or she does not struggle with diverging telephone numbers or e-mail addresses to prevent the customer from being left in loops forwarded from one employee or department to the other. This also involves not trying to get rid of customers' inquiries or problems by consoling the customer to another day.

"Horrible customer service. Was directed to different numbers/emails with no resolution for my issues."

After having received help and solution from the customer service, the customer should have the feeling that the retailer cares about their individual inquiries and problems. Individual contact to the retailer's employees via multiple communication media is vital for it.

"The self-checkout was a little intimidating until Eddie who works that area made it easy and pleasant experience"

R04: Prompt reply and solution

Similar to the problem solving ability, the customer service should solve the customers' inquiries and problems as quickly as possible. At best, at least a temporary solution should be given immediately after the problem is recognized and evaluated by the retailer's employee.

"Had problem just called and they had it fixed in 2 mins thanks"

Table 15: Implementation Guidelines and Service Design Requirements Overview

#	Guideline / Requirement	Description
Responsiveness		
R01	Continuous availability	The customer service should be available anytime a customer tries to contact it
G01	Provision of contact details	Provide contact information directly within the mobile app
R02	Multiple contact channels	Multiple channels should be available to contact customer service
G02	Integration of customer service chat	A chat should be integrated within the mobile app directly linked to a customer service employee
R03	Satisfaction with problem solving ability	The customer service should be able to assist the retailer's customers with any inquiries or problems
R04	Prompt reply and solution	The customer service should solve the customers' inquiries and problems as quickly as possible
R05	Politeness of customer service personnel	The customer service personnel should be polite and kind
G03	In-app guidance	The mobile app should include clear instructions for using the whole mobile app and single features
Information		
R06	Adequacy and completeness of information	The product and system information provided should be complete
G04	Exact sizing charts	Exact sizing charts should be provided for mobile clothing shopping
G05	Detailed update descriptions	Detailed update descriptions and release notes should be provided
R07	Clearness and understandability of information	The information provided through the app should be clear and understandable to the user
R08	Correctness of information	The information provided through the mobile app should be correct
G06	Correctness of stock information	The stock information should be displayed correctly for in-store and online shopping in the mobile app
Security and Privacy		
R09	Data protection against vulnerabilities	The customers' personal and payment data should be protected against external vulnerabilities
G07	Request password	The customer should log-in for using the mobile app
G08	Provide privacy policy	The retailer should provide transparent information on information processing and security
R10	Permission for data storage and processing	The customer's personal and payment data should only be stored if the explicit permission is granted
G09	Limited advertisement	Appropriate amount of advertisement in and caused by the mobile app
G10	Limited permissions for mobile app	The amount of permission requested by the mobile app should be appropriate
Design		
G11	Simple design	The mobile app should have a modern, clean design that fosters its intuitive use
G12	Clear and intuitive layout	The layout of the mobile app should be clean, simple and reduced
G13	High quality multimedia content	The mobile app should display high quality multimedia content aligned to the screen resolution
G14	Simple navigation	The navigation of the mobile app should be simple and convenient for the customer
G15	Reliable search engine	The search engine should provide the customer with the demanded results
G16	Convenient filters	Convenient filters should foster the search for requested items
Performance		
G17	Quick start-up	The mobile app should start up quickly without long waiting times for the customer
G18	Fast responsiveness	The mobile app should react responsively to the customer's interactions

(Continued)

Table 15—continued

G19	Adequate storage usage	The mobile app should occupy ass less storage as possible
R11	Reasonable network usage	The mobile app should cause as less network traffic as possible
G20	Reduced background activity	The background activity of the mobile app should be reduced
Technical Reliability		
R12	Accurate and consistent operation	The mobile app should be executed according to the description and promises given by the retailer
R13	Continuous availability of m-services	The m-services provided should be available anytime the customer wants to access them
G21	Sustainable updating	The mobile app should operate sustainably after the installation of an update
Valence		
R14	Overall satisfaction	The retailer should ensure that the mobile app satisfies the customers' needs and requirements
G22	Scope of features and services	The mobile app should incorporate all m-services provided by the retailer

Table 16: Responsiveness Design Requirements

Responsiveness	Implementation Guidelines / Service Design Requirements
RES1	R01: Continuous availability G01: Provide contact information directly within the mobile shopping companion app R02: Multiple contact channels G02: A chat should be integrated within the mobile shopping companion app directly linked to a customer service employee
RES2	R03: Satisfaction with problem solving ability R04: Prompt reply and solution
RES3	R05: Politeness of customer service personnel
RES4	G03: The mobile shopping companion app should include clear instructions for using the whole mobile app and single features

R05: Politeness of customer service personnel
Furthermore, the customer service personnel interacting with the customer personally or virtually should be polite and kind to make the customer feel treated well. Politeness should be ensured in all areas of contact to the customer either online or offline and also includes automatically generated documents such as electronic order confirmations. Experiencing a kind service will result in a higher MASQ.

"The customer service agent I spoke with was wonderful."

G03: In-app guidance
Considering the customer service within the mobile shopping companion app, it should guide the customer with additional information through the shopping process and provide information on all the features and services offered. There should be clear instructions available for using the mobile shopping companion app and all its features. An integrated manual should help the customer to get familiar with the mobile shopping companion app if it is not intuitive enough or the customer is not used to operate mobile apps.

"Can't find instructions on how to use, and there are features that are not all that intuitive IMHO"

5.2.2. Information

As the information quality dimension refers to quality aspects of information such as adequacy and correctness, the information design requirements introduce guidance on the provisioning of information to the customer through the mobile shopping companion app (tab. 17).

R06: Adequacy and completeness of information
The product and system information provided through the mobile shopping companion app have to be complete in the regard that the customer is able find every information he or she needs or is searching for at a glance. Specifically, customers request to receive information about recent offers and new items in the retailer's assortment.

"Used every day to keep updated with offers and developments"

As already mentioned in the information quality subsection, information not only comprises textual information and descriptions but also visual information using graphics, illustrations or even videos. Especially for non-food articles the

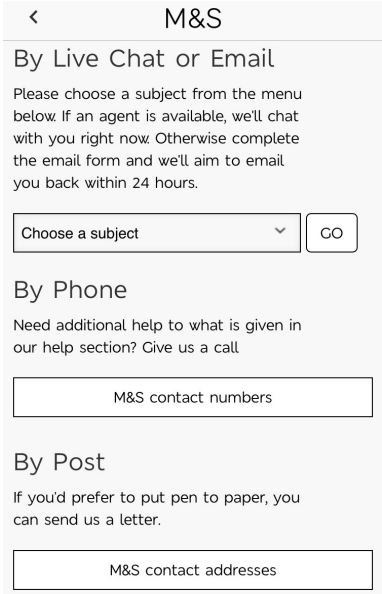


Figure 12: M&S App Contact Details Overview

Table 17: Information Design Requirements

Information	Implementation Guidelines / Service Design Requirements
INF1	R06: Adequacy and completeness of information G04: Exact sizing charts should be provided for mobile clothing shopping G05: Detailed update descriptions and release notes should be provided for each new release of the mobile app.
INF2	R07: Clearness and understandability of information
INF3	R08: Correctness of information G06: The stock information should be displayed correctly for in-store and online shopping in the mobile app.

visual information helps the customer to make an informed purchasing decision. So every item presented in the mobile shopping companion app should be accompanied by detailed visual information presenting every detail of the product supporting the customer's purchasing decision. This includes not only a single visual representation but images from different angles or even 360 degree views. "Everything else is pretty good with pictures of all items from multiple angles and clear descriptions."

Although a purchase can be processed completely within the mobile shopping companion app, customers require the mobile app to send confirmation messages via e-mail e.g. for processed orders or incoming items. They prefer to receive a confirmation of their mobile transaction in written form. These automatically generated messages also have to comply with the requirements mentioned for kindness and politeness.

"I think the confirmation of order message is excellent"

Information completeness also involves system information. An issue raised by customers using the Walmart app is that it requires a WLAN connection to provide all m-services. However, this information is not available to the customer, but only stating the m-services are not available at the moment a customer tries to access them using the mobile cellular network.

"To top it off, NOWHERE is this noted in your app or website documentation, and your associates have no clue, either. Every other app I have works whether on wifi OR cell data. What is wrong with your app????"

G04: Exact sizing charts

To justify the use of the mobile shopping companion apps, they have to deliver additional benefits to the customers compared to the use of the traditional website. Especially for the information quality dimension this means that the information available in the mobile app at least has to equal or better exceed the information that is available on the mobile website and in the store. This includes additional, detailed information about the product such as product characteris-

tics, ingredients or nutritional information and further reaching information such as receipt proposals or allergy alerts for certain ingredients based on the customer's predefined categories. As the virtual m-Commerce environment comprises certain limitations compared to the physical brick-and-mortar stores, the mobile shopping companion apps have to provide additional information to overcome these issues. One such issue is that purchasing through the e-Commerce or m-Commerce channel is the time-delayed handover of the item using a logistics service provider after the purchase compared to the brick-and-mortar stores in which customers directly take their purchases along. Another issue is that customers are not able to try on clothes. Thus, for shopping clothing in the appropriate size, the mobile shopping companion app should provide exact sizing charts as additional information. figure 13 shows the general sizing chart for women integrated in the M&S app. Under "Fit Guides" M&S also provides detailed instructions for the positions of measuring the different girths.

"And that is that there is no sizing charts for clothing. I can't make clothing purchases without one."

G05: Detailed update descriptions

Although mobile apps are frequently updated (Nayebi et al., 2016), developers should provide detailed update description and release notes every time a new version of the mobile shopping companion app is released. This enables interested customer to inform themselves about any changes made compared to the previous release and to decide whether they need or want to update or not. According to the customers, detailed update descriptions should include a list of bug fixes, added or removed features, security improvements and fixes against vulnerabilities as well as the complete date and version of the update. "When YOU update your app, you need to specify what changed features, bug fixes, etc. BS like WE MADE IT BETTER is not an update description."

R07: Clearness and understandability of information

To provide an additional benefit to the customer, information provided through the mobile shopping companion app should be clear and understandable for every user of the mobile app. The retailer has to ensure that the provided information is not misleading for the customer. A concrete manifestation of this design requirement is to provide consistent measurement units for product size and weight at least throughout a product category. This allows customers to easier compare items and make a more informed purchase decision.

"Shame you can't compare times as some are given in pence per kg and others pence per gram, making comparisons difficult."

R08: Correctness of information

The information provided by the mobile shopping companion app and related m-services should be correct. Correctness of item information in this case means that the product information resembles the real information about the product. Attaching the items to the adequate product categories is also part of information correctness. An important characteristic of correct information is its topicality. Correct information always has to be up to date. In an optimum manner, this information should be updated in real-time if it is changed in store or in the retailer's back-end systems. Further, the graphical representation of the products in the mobile shopping companion app should match reality and presented characteristics of the product (e.g. blue trousers should be displayed in blue). In addition, price information has to be correct (minor differences compared to the in-store prices are possible). In-store and online prices rarely equal each other and often times the prices differ from store to store.

"I bought a lawnmower that on the website clearly shows it was self propelled. I get to the store and it wasn't."

Correctness of information is also considered with system information related to technical specifics of the mobile shopping companion app and purchase process information such as delivery tracking and other notifications.

"Don't understand why I'm not getting notifications when my order is ready for pickup.??" An example of the importance of correct information is given by a user of the Walmart app. For a certain area of the United States, the store locator feature also displays distribution centers as stores for customers. However, when using the store finder feature, the mobile shopping companion app should only show brick-and-mortar stores and no distribution centers as customers usually cannot purchase anything there.

"If someone is using the app and looking for the store nearest to them, why would you have a distribution center come up as the closest store and then direct people there? Only actual stores should come up when looking for stores"

G06: Correctness of stock information

Another manifestation of the importance of correct information is the correctness of stock information that is provided through the mobile shopping companion app. It should provide accurate and correct information on the availability of items and the actual number of items available in each brick-and-mortar store and for the online shop. The in-store stock availability information is important to mention here, because customers tend to rely on this information and go to the store assuming that the product is available and get dissatisfied if it is not and the stock information provided by the mobile app appears to be false. By analyzing the online customer reviews in the QCA it seems as if some stores do not update the stock information immediately leading to the described discrepancies in the stock availability.

"I downloaded this app to see if an item was in Stock at a store. It says it was. I drove all the way there. It was not there. What good is an app that says something is in Stock if it is inaccurate?"

5.2.3. Security and Privacy

Security and privacy are very important in the m-Commerce context with less control over private and payment data (Kuo et al., 2016, p. 5). The security and privacy dimension of the MASQ measurement scale is concerned with the protection of the customers' private and payment information. Thus,

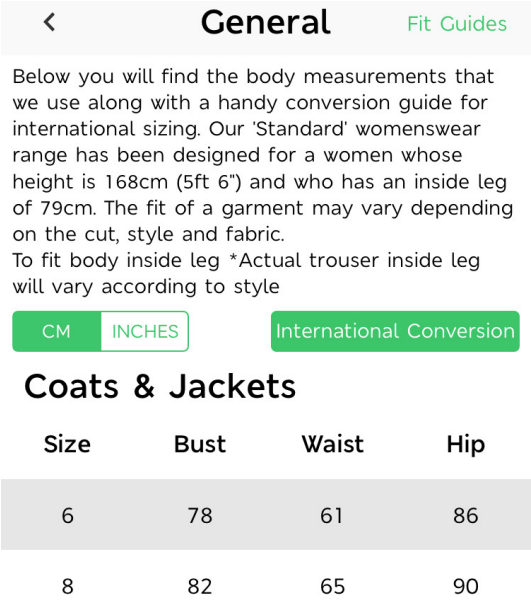


Figure 13: M&S General Sizing Chart

the design requirements and implementation guidelines presented are mainly concerned with procedures and methods protecting any personal information (tab. 18).

R09: Data protection against vulnerabilities

Because personal customer and payment data is a valuable resource in the information age, the retailer has to ensure that personal customer and payment data is not leaked to unauthorized people neither through external hacker attacks nor internal data leaks (Eaton and Bawden, 1991). If the data protection is not successful at a certain time and a data leak occurs, the retailer should reduce the damage for the customer and offer compensation for the losses.

"Someone hacked my account and took $ 35.00 from me and Walmart didn't do anything about it!"

The analysis of the online customer reviews reveals several security features requested or even taken for granted by the customers. Firstly, the customers' personal and payment data should be encrypted on the mobile device, on the data transfer via the mobile internet and on the retailer's back-end systems.

G07: Request password

Secondly, a way to protect the customer's personal and payment data on the mobile device is to request for a sign-in every single time the customer wants to use the mobile shopping companion app although some customers complain that it is annoying to re-type the password each time using the mobile app and at important phases of the purchasing process such as the check-out. The password required for signing-in should be as complex as possible[11] and not be restricted to a short length of twelve characters as limited by Walmart on its Android app. This customer complaint can be weakened by integrating fingerprint-based log-in mechanisms such as Apple TouchID simplifying the log-in procedure for the customer.

"Forcing re-entry of password for checkout is annoying."

G08: Provide privacy policy

Thirdly, further customer complaints and doubts especially on providing and storing credit card information can be addressed by providing detailed and transparent information regarding the data security, data storage and usage of personal and payment data (Rudolph et al., 2015, p. 47). The three analyzed mobile shopping companion apps include the retailers' customer privacy policy statements. Customers can get detailed information on the information security and data protection. The retailers highlight their measures to secure and protect the customer' personal and payment data. Moreover, they explain and justify the data processing and collection through the mobile apps. An exemplary part from the customer privacy policy provided by M&S is presented in figure 14. Trust in the usage of mobile shopping companion apps can be further increased by applying renowned quality labels (Rudolph et al., 2015, p. 48).

"Oh and check the privacy policy by following the link on this page and you get 'Sorry, the page you are trying to access does not exist.'"

R10: Permission for data storage and processing

For storing and processing the customer's personal and payment information, the customer's explicit permission is required. Although it seems to make the purchasing process more convenient for the customer, storing his or her personal or payment data should only be done after the customer has explicitly granted the permission for storing his or her data. This also applies for tracking, storing and processing the cus-

[11] The National Institute of Standards and Technology regularly releases guidelines concerned with the complexity of passwords.

Table 18: Security and Privacy Design Requirements

Security and Privacy	Implementation Guidelines / Service Design Requirements
INF1	R09: Data protection against vulnerabilities G07: The customer should log-in for using the mobile app. G08: The retailer should provide transparent information on information processing and security.
INF2	R10: Permission for data storage and processing G09: Appropriate amount of advertisement in and caused by the mobile app.
INF3	G10: Only request permissions required for accessing the m-services provided by the mobile shopping companion app.

tomers' shopping habits online and in the brick-and-mortar stores.

"Major privacy concern I just used the app to make a purchase. Without my knowledge or permission, it stored my credit card information. This is absolutely unacceptable!"

G09: Limited advertisement

The mobile shopping companion app should not overload the customer with advertisements. The amount of advertisement displayed in and caused by the mobile app has to be appropriate. The advertisements are usually integrated in the mobile app in the form of banners, notifications or images but the mobile shopping companion app can also cause advertisement emails. For providing advertisements to the customer, the customer should be asked for the explicit permission. Even if this permission is granted, the customer should not be overloaded with advertisements and not distracted from the actual purchasing process.

"Way too many ads. So much so that I could not complete a purchase and had to finish on my computer."

G10: Limited permissions for mobile app

The mobile shopping companion app should only request permissions that are necessarily required for the proper provision of its m-services and features. Permissions to access personal information and smartphone functions not required for providing the m-services seems not to be appropriate at first glance. It is not the mere number of permissions required but also their quality with regards to the interference of private information and the possibility to cause additional costs for the customer. Furthermore, the mobile shopping companion app should not grant itself permissions the customer has not granted before or the customer has intentionally removed. In this regard, the mobile shopping companion app should be executable even if not all permissions are granted. At least these features not requiring permission should operate properly.

"Unwarranted permissions - why does it need access to WiFi info, media files and photos, contacts - no, really M&S, and no explanation either Again: why these dangerous permissions."

5.3. Environment Design Requirements

Environment design requirements subsume the design requirements identified in the QCA that refer to the environment quality primary dimension. Because this section is concerned with the Design and Performance of the mobile shopping companion apps, it mainly provides implementation guidelines as suggestions for improving the MASQ of these mobile apps.

5.3.1. Design

In line with the description of the design SQ dimension as presented in subsection 4.3.1, design implementation guidelines for this dimension include suggestions for the aesthetic aspect of the UI and filter and search function (tab. 19). These suggestions are tailored for the characteristics of mobile devices such as relatively small displays and touch control.

G11: Simple design

The UI design of the mobile shopping companion app should be modern and clean to foster the intuitive use of the mobile app. It should follow a flat and straight design approach involving color contrast to ensure a better clarity for the customer (e.g. fig. 7, 10). To overcome the restrictions regarding the display size, developers should design the usability of the mobile apps as easy and intuitive as possible. Operating and navigating through the mobile apps should be simplified by using symbols and pictures instead of huge amount of textual information. However, complex and large graphics should be omitted to avoid long loading times (Rudolph et al., 2015, p. 47).

"Very simplistic user interface which could be more intuitive."

Although it is difficult to fulfill all the customers' individual preferences with regards to the used color scheme, the used colors should be appealing to the majority of customers. In addition, it should fit the corporate identity of the retailer and transfer the retailer's vision to some extent.

"Good size text and clear with suitably neutral colours which are pleasing on the eye."

G12: Clear and intuitive layout

The layout of the mobile shopping companion app should be

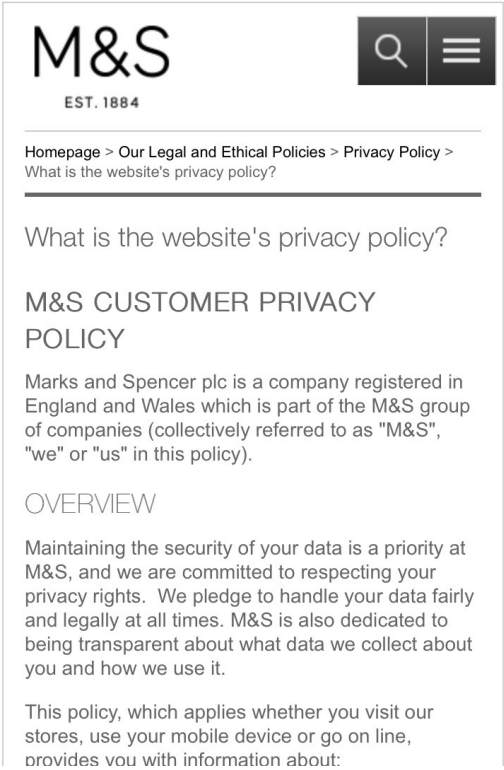


Figure 14: M&S App Customer Privacy Policy

Table 19: Design Design Requirements

Design	Implementation Guidelines
DES1	G11: Mobile shopping companion apps should have a modern, clean design that fosters its intuitive use. G12: The layout of the mobile shopping companion app should be clean, simple and reduced.
DES2	G13: Mobile shopping companion apps should display high quality multimedia content aligned to the screen resolution.
DES3	G14: The navigation of the mobile app should be simple and convenient for the customer.
DES4	G15: The search engine should provide the customer with the demanded results. G16: Convenient filters should foster the search for requested items.

clean and simple. It should be reduced to essential elements directing the customer's focus on the items and features provided. Having implemented a layout, it should be consistent through every single page of the mobile shopping companion app. A consistent layout is more intuitive for the customers because a uniform structure of the UI does not require learning and understanding multiple layouts.

"Love the way items are arranged in the categories for pickup discounts and price rollback"

For the customers one aspect of a clear layout is the spatial separation of UI elements. Overlapping UI elements not only may distract the customer when using the mobile shopping companion app but also may make information of the rear objects invisible and unreadable for the customer.

"The green purchased checkmarks on your list are annoying because they cover the item picture."

G13: High quality multimedia content

As mobile shopping companion apps include various kinds of multimedia content such as icons, graphics, sounds and videos, this content also should fulfill certain quality characteristics with regards to resolution and graphical quality. Moreover, the quality of the multimedia content should be appropriate to and fit the screen size and resolution. The image and video resolution should be high enough to view

them on a mobile device in full screen. With the introduction of mobile devices with increasingly larger screen sizes and higher display resolutions, not only the multimedia content has to be available in higher resolution but also the whole mobile shopping companion app has to be tailored to the new resolution.

"doesn't get 5 stars because of the ui and graphics. Low resolution or not optimized for retina screens."

Especially graphical product representations should be of high quality and resolution because pictures of the product may influence the customer's purchasing decision. Thus, the representation of the products should be close to reality presenting all necessary product details to the customer. The product representation should adapt according to the selection made by the customer with regards to e.g. color, size or cutting. Moreover, the pictures should be enlargeable to make more product details visible for the customer. In this regard, it is useful to offer multiple pictures from multiple angles to give the customer a complete overview. Optimally, these multiple pictures should build up a 360 degree view which is controllable by the customer.

"Pictures are clear and give good detail of the product."

G14: Simple navigation

An easy use and navigation is the top mentioned requirement by the users of the three mobile shopping companion apps under investigation (tab. 14). Navigation and maneuvering through the mobile shopping companion app should be as convenient as possible for the customer. Navigation should be intuitive not requiring long time periods for learning. Although it is certainly meant ironically, some customers complain that the mobile app is "Too easy to use" because using the mobile app it is simple and fast to spend more money. The mobile shopping companion app should involve a flat menu structure enabling the customers to find items without the necessity of navigating through many hierarchical menu levels.

"This app is very easy to navigate and find your way round"

Commands and buttons that are important to the customer and often used should be easy accessible. Among others, these are a button for signing out of the mobile app and the final check-out button. Analogous to the consistency of the layout, the navigation should be similar through the whole mobile app.

"Navigating around the pages for shopping is fine, but difficult to navigate around account & Sparks information and contacting M&S with queries."

The mobile shopping companion app should offer an easier access to the retailer's online shop on the mobile device than the mobile version of the website. The mobile app should be more responsive and easier to navigate making use of advanced technologies of the mobile device. "This app is far quicker & easier to use than the website."

G15: Reliable search engine

The mobile shopping companion app should integrate a reliable search engine to query the assortment and other features of the mobile app. The search engine should provide the results the customer is searching for with a particular query. Thus, the search result has to match the search query. To allow for an appropriate search result, the search engine should allow the customer to use synonyms and paraphrases in case the exact denotation of an item or brand is unknown.

"Also a search for 'high heeled shoes' only gave 2 results whereas searching for Shoes and filtering on women's and High heel height brought up a whole load of shoes"

To enable a quick product search, the search function should be accessible from everywhere in the mobile app. In addition, the search engine should not only search the retailer's assortment but also other features and services of the mobile shopping companion app. For instance, it should be able to search previous orders for particular items or search for the customer's loyalty card identification number.

"Also, the search function is only available on the home screen, that's really not practical."

G16: Convenient filters

Besides the presented search engine, retailers should implement the possibility to filter search results and even whole product categories in their mobile shopping companion apps as additional instrument for identifying the product desired by the customer. The filters should allow the customer to narrow down the search results or the whole assortment as detailed as possible. Further, the number of filters and the product characteristics that can be filtered should be convenient for the customer. Customers are used to apply filters because they are usually integrated in web-based shops. Thus, the filter functionality of the mobile shopping companion app has to meet the functions provided on the website otherwise it appears as providing a lower SQ. This includes the possibility to apply multiple filters at the same time and even to build up a filter hierarchy.

"The filter is excellent, no need to waste time looking through items not available in your size or the colour you want etc."

Filtering products and search results should also include on option to order the products according to certain characteristics (e.g. price or size). This additional function of the filter enables customers to identify the desired product even faster.

"Why doesn't this app allow me to sort my results? Like price high - low?"

5.3.2. Performance

Performance design requirements deal with suggestions to improve the overall responsiveness of the mobile shopping companion app also containing issues with the mobile internet connection in brick-and-mortar stores and performance-related aspects of the mobile device executing the mobile app (tab. 20).

G17: Quick start-up

Smartphone users are accustomed to an almost immediate access to their device and their installed mobile apps. Hence, mobile shopping companion apps should also start-

Table 20: Performance Design Requirements

Performance	Implementation Guidelines / Service Design Requirements
PERF1	G17: The mobile app should start up quickly without long waiting times for the customer. G18: The mobile app should react responsively to the customer's interactions.
PERF2	G19: The mobile app should occupy ass less storage as possible. R11: Reasonable network usage
PERF3	G20: The background activity of the mobile app should be reduced.

up as quickly as possible without causing long waiting times for the customers. At best, it should be available immediately after the customer has opened it. The customer should not get the impression that he or she has to wait for the mobile shopping companion app to start-up. The quick start-up also involves a fast log-in procedure that might be necessary for reasons of security. "Opening app takes a long time."

G18: Fast responsiveness

After a quick start-up, the mobile shopping companion app should also react responsively to the customer's input enabling an interruption-free interaction between the customer and the mobile app. The mobile shopping companion app should respond to the customer's interactions almost immediately. The customer should not get the impression that he or she has to wait for the mobile app to respond. This design requirement does not involve waiting time that is caused by a slow connection to the internet induced by an insufficient WLAN or mobile cellular connection. The responsiveness should not be influenced by the number of customers accessing the m-services at the same time. In case the mobile shopping companion app still freezes or lags it should continue its operation after an appropriate amount of time and should not crash completely requiring a restart because a restart always involves the danger of losing previously entered information.

"It takes 20 seconds just to add or remove items from my cart and when it was time to check out it took me 10 minutes to go through 3 simple steps"

G19: Adequate storage usage

The overall storage capacity of mobile devices is limited. Even the increase of the storage capacity of mobile devices over the last years does not solve this issue since mobile apps, the mobile operating system and multimedia content also require more storage capacity (Stahlknecht and Hasenkamp, 2005, p. 351). Therefore, the mobile shopping companion app should only use as less storage of the mobile device as required for providing all of its features and m-services. A mobile app requiring much storage while not providing enough benefit is not worth being installed for some customers. Because the internal storage of mobile devices is limited and part of it is already occupied by the operating system of the mobile device, some of them allow attaching of an additional memory card to extend the storage capacity. To clear storage capacity on the internal storage for other purposes, the mobile shopping companion app should be transferable to the additional memory card if available.

"With storage space at a premium on my phone, I am seriously reconsidering the necessity of having this app."

R11: Reasonable network usage

Even though the overall mobile network usage and volume of data per mobile line is steadily increasing, the aim of each mobile app should be to consume less mobile network volume to keep it for other purposes of the user (Bundesnetzagentur, 2017; VATM, 2017). Hence, the mobile shopping companion app should use as less mobile network data as required for providing its features and m-services. Especially, avoidable network usage should not be generated. Avoidable network usage in this case are bytes transferred via the mobile cellular network not necessary for providing the features and m-services of the mobile app (e.g. advertisement including large graphics). Moreover, the mobile shopping companion app should be able to access the m-services using both the mobile cellular network and WLAN. Although this should be an inherent feature of mobile apps, several customers reported that some features of the Walmart app only seem to work with mobile internet access via WLAN. Furthermore, due to sometimes complex structures of brick-and-mortar stores, the connection to the mobile cellular network can even be slow with potential dropouts. Therefore, retailers should set-up a free WLAN to decrease the loading times and increase the experience for using the mobile shopping companion app (Rudolph et al., 2015, p. 47).

"Too data intensive. I am simply trying to refill a prescription (my ONLY use for this app) and it is taking FOREVER to load a bunch of ads that are rotating and won't let me go to the pharmacy link"

G20: Reduced background activity

Recent mobile apps perform some activities in the background even when the mobile app is not running in the foreground or the mobile device is locked. These background activities consume resources of the mobile device such as processing power or network connection that will require battery power leading to a reduced durability of the battery. Although some background activities are required to keep the mobile shopping companion app updated with current information for the customer, others are avoidable and only "wasting the battery" (Ravindranath et al., 2012, p. 118). Therefore, mobile shopping companion apps should only be

executed in the background if it is necessary for the provision of the m-services and then go easy on the resources of the smartphone. “It is active in the background to such a degree that it is a constant drain on my battery– and constant activity means increased and continual risk to my privacy and safety.”

5.4. Outcome Design Requirements

The outcome design requirements integrate the implementation guidelines and design requirements identified during the QCA that reference to the outcome quality primary dimension of the MASQ measurement scale. They are further subdivided according to the related secondary dimensions Technical Reliability and Valence.

5.4.1. Technical Reliability

Technical reliability design requirements are concerned with measures ensuring an accurate and consistent performance of the mobile shopping companion app. The mobile app should be executed without any failure on the mobile device from a technical point of view (tab. 21).

R12: Accurate and consistent operation

The mobile shopping companion app, its features and provided m-services should be executed accurately and consistently according to the description and promises given by the retailer. Although developers frequently provide new releases of their mobile apps (Nayebi et al., 2016), they have to ensure a certain quality for each release. The QCA reveals that many online customer reviews deal with technical malfunctions of the mobile app as a whole or single features. Customers not only report the malfunctions but also express their disappointment that may lead customers to stop using the retailer’s mobile shopping companion apps. Even though a single feature does not work or an m-service is not available, the mobile shopping companion app should not crash.

“Nullpointer exception when trying to add anything to basket”

For several malfunctions, the reasons for their occurrence are not obvious. Nevertheless, the developers should monitor the customers reports about failures and malfunctions and fix them as quickly as possible within a new release of the mobile app. According to the QCA, the Walmart app has issues with the barcode scanning feature within the savings catcher. As customers already report some problems with the camera permissions, the developers should provide a workaround until the malfunction is fixed in a new release.

“Can’t get scanner on App to work. When scan is selected on search you just get a black screen"

R13: Continuous availability of m-services

Mobile shopping companion apps provide mobile access to services such as the online store that are executed and hosted on the back-end systems operated by the retailer. The m-services provided by the mobile shopping companion app should be available anytime the customer wants to access them. Thus, they should be available 24 hours a day. Once customers get used to the mobile provisioning of these services, they rely on using them. Hence, an unavailable m-service may negatively influence the overall perceived MASQ. In case an m-service is temporarily not available, the retailer should ensure that it will continue operating as soon as possible and should apologize for any inconveniences.

“Keep on getting message. Oops we are having temporary system issues... Simply doesn’t work!”

The mobile shopping companion apps are intertwined with other mobile apps of the retailer and also with third-party m-services like Apple Pay[12]. Thus, it is difficult to determine the cause for a malfunction of the mobile app. It could also be a problem with the referenced mobile app, m-service or the interface between them. However, customers tend to charge the malfunction to the mobile shopping companion app itself. Only in case of the involvement of third-party m-services they acknowledge other causes.

"My checkout using Apple Pay failed once but went through the second try. Not sure who is to blame here Apple or M&S."

G21: Sustainable updating

As the mobile apps are frequently updated (Nayebi et al., 2016), the retailer should guarantee that the mobile shopping companion app operates sustainably after the installation of an update as it has done previous to the update. Developers have to ensure that the transition to a new release of the mobile app is smooth for the customer. At best, the customer does not have to take any additional effort to make the mobile shopping companion app operate accurately and consistently after the installation of a new update. The mobile shopping companion app should also remember the customers’ personal and log-in information if they are stored before. In addition, new releases should introduce technical, aesthetical or functional improvements to the customers and most importantly fix known issues.

“STILL can’t use scanner...ever since November 2016 update. Walmart WHEN are you going to fix the camera interface?”

Moreover, the mobile shopping companion app should give correct notifications if an update is required. It should not mislead the customer with a notification to a new release if the installed release of the mobile app is already the most current version. Several customers of the Tesco app reported in the online customer reviews that they are notified by the mobile shopping companion app that it needs to be updated although the most current release is already installed on the mobile device. Unfortunately, the mobile shopping companion app cannot be used if the notification for the update appears. When the customer presses the blue "OK" button, the mobile shopping companion app is closed and the customer is redirected to the MADP whether a new release of the mobile app is available for installing or not. The update notifi-

[12] Apple Pay is an m-service introduced by Apple Inc. allowing users to pay securely online and in-store using existing contact-less payment terminals that is only available on recent mobile devices designed by Apple.

Table 21: Technical Reliability Design Requirements

Technical Reliability	Implementation Guidelines / Service Design Requirements
REL1	R12: Accurate and consistent operation
REL2	R13: Continuous availability of m-services
REL3	G21: The mobile shopping companion app should operate sustainably after the installation of an update.

cation nearly covering the whole screen of the mobile device is depicted in figure 15.

“Caught in an infinite loop, tells me app is out of date even though it’s just been installed. Useless”

5.4.2. Valence

The valence design requirements deal with the customer’s satisfaction when using a mobile shopping companion app and highlight the necessity of a satisfied customer after having used the mobile app. The valence quality dimension is one of the most important quality dimensions in the MASQ scale. It summarizes the customer’s overall feelings and perceptions about the mobile shopping companion app and the m-service provided from the information to the after-sales phase (tab. 22).

R31: Overall satisfaction

By fulfilling the previous implementation guidelines and service design requirements, the retailer should ensure that the mobile shopping companion app satisfies the customers’ needs and requirements. After the customer completed a transaction he or she should be satisfied by an enhanced SQ perception compared to the traditional and electronic service environment. Besides the functional and technical satisfaction, the customer should also be emotionally satisfied by aesthetical aspects such as color or appealing sounds. “Honestly, it’s the worst shopping app I’ve ever tried to use. Truly awful in every respect.” Furthermore, the QCA reveals that the customer’s quality perception of the mobile shopping companion app often replaces the quality perception of the brick-and-mortar store itself and even the whole company when shopping with mobile devices. In extreme cases, this can lead customers to shop at other retailers if the perceived MASQ of the mobile shopping companion app is unsatisfactory and thus can negatively influence the retailer’s revenues.

"The last few weeks have had nothing but problems. Today it showed no available slots for 7 days. What’s that about. Thinking of going to Sainsburys!"

Inversely, the expected SQ of the mobile shopping companion app is influenced by the perceived SQ in the traditional brick-and-mortar stores. A high SQ in the store will lead customers to expect an equally high MASQ for the mobile shopping companion app. However, the dimensions for measuring the two qualities are different.

"Overall a very poor experience which is a shame for a good company like M&S" Moreover, a satisfied customer is likely to recommend the mobile shopping companion app to other potential customers such as friends and relatives. This traditional word of mouth may lead to more customers using the mobile app and may also positively influence the retailer’s revenues. “would defiantly recommend this to friends”

While a number of customers is dissatisfied about the security and privacy issues related to the increasing dissemination of mobile apps in further spheres of life such as the in-store and online shopping, others clearly express their satisfaction with a digitized environment. Hence, retailers should further strive for entering their customers’ everyday routines to further improve their shopping experience and loyalty with the retailer. To foster the dissemination of their mobile shopping companion apps, retailers have to try to reduce the barriers for using them. These are among others technical barriers such as relatively small displays of the mobile devices and slow connection to the mobile internet and concerns about the security and privacy of mobile devices and mobile apps (Rudolph et al., 2015, p. 47).

"Thankfully, all that has changed in the world we live in, full of technology. What better way to know exactly where you spend your money, what on, and how to better tailor future advertisements specifically to you!"

G22: Scope of features and services

Analyzing the three mobile shopping companion apps, retailers seem to follow two different strategies when implementing their mobile apps. Either they implement one mobile shopping companion app integrating all features and m-services the retailer provides or they distribute a number of mobile apps dedicated to a single or a small set of features and m-services. While the former approach is used by Walmart with its "Walmart: In-Store & Online Shopping. Easy Reorders" app, the latter is followed by Tesco. Besides a small set of mobile apps that are tailored for a specific target group, the Walmart app includes all the features and m-services provided by Walmart (app. B). An overview of the features and m-services provided by the mobile shopping companion app of Walmart is given on the start screen. This overview is shown in figure 16.

The "Groceries" app provided by Tesco is only one of many m-Commerce-related mobile apps. The Tesco app includes references to other m-Commerce-related mobile apps, m- services and the mobile website (fig. 17). The "Groceries" app contains only a limited set of features (app. B).

“You have so many apps doing different things! Payqwiq,

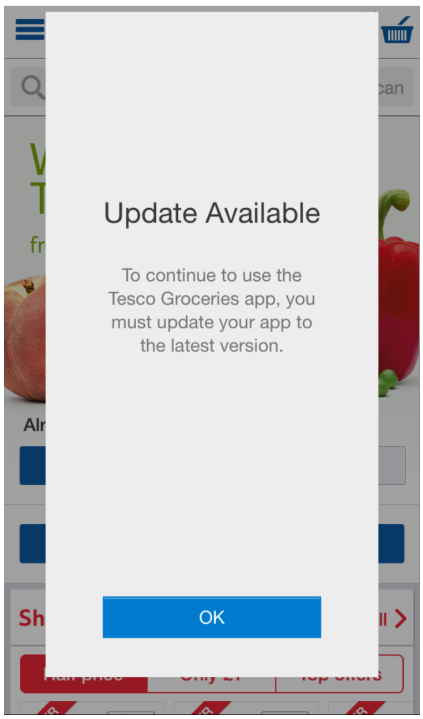


Figure 15: Tesco App Update Notification

Table 22: Valence Design Requirements

Valence	Implementation Guidelines / Service Design Requirements
VAL1	R14: Overall satisfaction
VAL2	G22: The mobile shopping companion app should incorporate all m-services provided by the retailer.

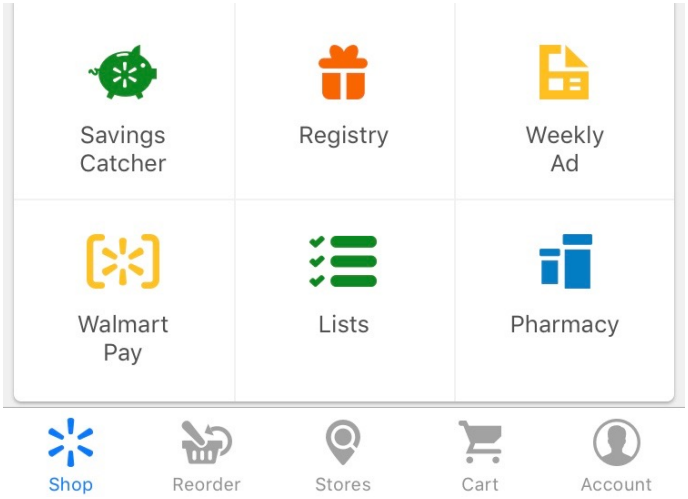


Figure 16: Walmart App Scope of Internal Services

Clubcard and Groceries. Link them for easy use."

The online customer reviews concerned with this topic, suggest that customers prefer to implement mobile shopping companion apps according the first approach having a single mobile shopping companion app encapsulating all features and m-services. Following this approach would prevent customers from constantly switching between different mobile apps e.g. for completing a purchase transaction. Further, developers do not have to take care of links and interfaces between these intertwined mobile apps and can focus on the proper integration of third-party m-services like Apple Pay.

"Everything you need in one place, Sparks and online shopping."

The scope of features and m-services covered with mobile shopping companion apps is further emphasized in section 5.5.

5.5. Analysis of Features

As nearly 400 online customer reviews deal with the specific features already integrated in mobile shopping companion apps or additional features desired by the customers, in the following section an analysis of features of mobile shopping companion apps will be provided along with a proposition for a set of features based on the QCA.

5.5.1. Diversity of Features

Customers demand a certain set of features and services to be likely to rate a mobile shopping companion app as useful. This set of features has to solve the possible problems of the customer (Madu and Madu, 2002, p. 251). Therefore, it is important which "other features are available" (Madu and Madu, 2002, p. 251) besides the mobile access to the online store that constitute a mobile shopping companion app. Additionally, the implementation quality and the resulting usefulness for the customer is a determining factor for the diversity of features demanded by the customers.

A first indication of the features and m-services provided by the mobile apps is already given with the definition of mobile shopping companion apps (subsection 2.3.2). Because the features dimension is only recognized as determinant of ESQ by Madu and Madu (2002) and Collier and Bienstock (2006), it is excluded from the actual MASQ measurement scale. Additionally, a possible features dimensions could not be measured by analyzing the mobile shopping companion app itself but only by interviewing customers about their expected and desired features and m-services.

As the QCA reveals, there are several online customer reviews in which the customers express recommendations for additional features that should be added to the mobile shopping companion app. The additional features should tailor the mobile shopping companion app to the specific problems and needs of the customers. Often, these features are familiar to the customer as they are already implemented in other mobile apps they use more or less frequently. And they consider this particular feature as useful as it has to be offered by the mobile shopping companion app as well. More rarely, customers want the developers to re-implement features that have been removed from the mobile shopping companion app because of some reason.

In addition to the recommendations, other online customer reviews incorporate criticism and praise for implemented features. This feedback allows to deduce if these features should be integrated in a mobile shopping companion app.

"I am loving the Apple Pay option at checkout."

5.5.2. Proposed Set of Features

This subsection should involve concrete suggestions on the composition of the set of features provided by the mobile shopping companion app. In this regard, only features and m-services are presented in this subsection that are mentioned and requested by a majority of customers in the QCA. Moreover, the features introduced resemble to a high degree the m-services suggested by Rudolph et al. (2015) in subsection 2.3.2 such as online purchase, shopping lists or loyalty card manager. The exact set of features and m-services provided by a mobile shopping companion app has to be defined by every retailer based on their customer group's demands and desires (Rudolph et al., 2015, p. 43). Table 23 presents an overview of the set of features including a short description that an exemplary mobile shopping companion app should provide.

Mobile app dissemination

During the SLR some features were identified that allow a larger group of customers to use the mobile apps. Firstly, multilingual support should be implemented. This is especially necessary for retailers operating internationally. Secondly, facilitated conditions for disabled people are required to enable them to also access the m-services such as the possibility to enlarge the font size or change the color scheme (Collier and Bienstock, 2006, p. 264). Furthermore, some customers requested a voice-over reading out the textual descriptions.

"Using Voiceover: You can't double tap on an item to get its further details."

Seamless integration into brick-and-mortar environment

Mobile shopping companion apps are defined as digital companions supporting the shopping and enhancing the customers' shopping experience within and outside the traditional brick-and-mortar stores using mobile technology (sec. 2.3.2). To support the shopping within a traditional brick-and-mortar store, the mobile shopping companion app and its provided m-services have to be integrated in the traditional environment conveniently for the customer. The customers should not experience any disruption or transition using an m-service provided by the mobile shopping companion app while doing the shopping in a brick-and-mortar store. Therefore, they have to be fully integrated and harmonized with the in-store equipment and personnel. The in-store interaction between the customer and the retailer represented either by its employees or equipment should be improved by using mobile shopping companion apps. Although the three

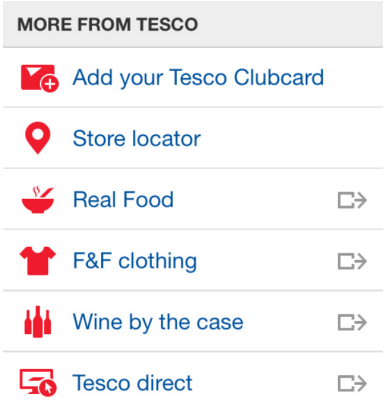


Figure 17: Tesco App References to External Services

Table 23: Proposed Set of Features

#	Feature	Description
Mobile app dissemination		
F01	Multilingual support	Integrate multiple languages into the mobile app
F02	Facilitated conditions	Offer features for easier access for disabled people
Seamless integration into brick-and-mortar environment		
F03	Shelf-locator	Provide the in-store location of an item
F04	Loyalty card scanning	Include a scannable representation of the digital loyalty card on the mobile device
F05	Stock availability	Check in-store and online product availability
F06	Mobile wallet support	Make electronic cards available without launching the mobile app
F07	Mobile payment services	Integrate popular mobile payment services
F08	Savings catcher	Guarantee lowest price for customers based on competitors and other stores
Improving the mobile shopping experience		
F09	Landscape mode	Change orientation when mobile devices is turned into horizontal position
F10	Cross-device linkage	Exchange information across devices based on customer accounts
F11	Online purchase	Provide access to the online shop
F12	Favorites list	Save customer's favorite items on a special list
F13	Continuous basket	Store customer's basket if the mobile app crashes
F14	Item amendment	Amend items to a purchase until final delivery
F15	Swipe pictures	Swipe through multiple pictures
F16	Loyalty card manager	Manage the complete loyalty account on the mobile device

mobile shopping companion apps by Walmart, Tesco and M&S integrate some features for improving the customers' in-store shopping experience, these features can be further enhanced.

F03: Shelf-locator

The mobile shopping companion apps by Walmart and Tesco offer an aisle-or shelf-locator to simplify the in-store product search by providing the aisle number at which the product is located. However, this oftentimes cryptic number is not helpful if it is not displayed in the store. For an enhanced shopping experience, it would be even better to directly include a map of the brick-and-mortar store showing the exact location of the product the customer is looking for. Thinking further the idea of the shelf-locator combined with a map of the brick-and-mortar store, can result in an in-store navigation taking into account the customer's position identified using the location-awareness of the smartphone. By doing so, customers could be guided through the brick-and-mortar store according to their predefined shopping list within the mobile shopping companion app or their personal shopping

behavior and interests.

"Thank you for letting me know that item is on L17. But where the f*** is L section. Useless information. What is good to know that it is on L17 when I can't find f***ing L."

F04: Loyalty card scanning

At M&S the cashiers have to type in the identification number of the electronic loyalty card, called Sparks, manually at each check out. It would reduce waiting times at the check-out if the mobile shopping companion app would provide a barcode, quick response code or another visualization that can be simply scanned from the display of the smartphone at check out. Furthermore, recent smartphones and Point of Sales (POS) systems support Near Field Communication (NFC) for the wireless exchange of information over short distances. Thus, the information of the digital loyalty and credits cards can be exchanged wirelessly at the POS using NFC.

"However many branches I go to cannot scan the barcode and the poor assistant has to type the massive number in"

F05: Stock availability

The mobile shopping companion app should provide the customer with information regarding the actual number of items in each store and available for online purchase. Optimally, this information should be updated in real time as customers usually rely on this information when going to a local store for purchasing a particular item. In addition, there should be a hint if the rest of the shelf is below a certain limit to indicate that the item may be out of stock when the customer arrives.

Annoying when everyday items display as not available, when you then go into the store to find hundreds available on the shelf"

F06: Mobile wallet support

Currently, the customer always has to log-in to the mobile shopping companion app to retrieve his or her loyalty card or payment information at check-out. As the process of logging-in may take some time due to e.g. slow mobile internet connections in the brick-and-mortar stores, the shopping experience can be enhanced by implementing support for Apple or Android Wallet[13]. Supporting the mobile wallet apps allows customers to access their loyalty card or payment information without an additional log-in to the retailer's mobile app. This additional feature further accelerates the check-out process at the POS. "Where is the Apple wallet support? I can't find anywhere allowing me to access my Sparks card from the lock screen via the Wallet. Have to do it from the app, which is a big hassle."

F07: Mobile payment services

To improve the check-out process in-store and for mobile shopping, the retailers often integrate their own proprietary payment options to their mobile shopping companion apps.

However, customers complain about requiring an additional payment, credit card or account just for checking-out for mobile transactions. Hence, retailers should implement third-party payment options in their mobile shopping companion apps such as Apple or Samsung Pay to enable customers to reuse their existing accounts without the need to create new accounts or payment options for m-Commerce. These mobile payment services even offer security mechanisms modified for the mobile environment.

"Great App. Simple to use, but would be so much better with Apple Pay integration for faster and secure payments."

F08: Savings catcher

However, the savings catcher feature offered by the Walmart app is introduced as unique selling proposition in subsection 5.1.1, it is often rated as a useful feature and main reason for using Walmart's mobile shopping companion app in the online customer reviews. It provides customers with discounts in form of an electronic gift card for products that are cheaper at other stores of the retailer, in the retailer's online store or at a local competitor. By using the savings catcher, the customer usually purchases for the lowest available price. Hence, implementing a similar feature could be worthwhile for other retailers. "savings catcher, finds the lowest price locally and refunds me the difference if I paid more at thier store."

Improving the mobile shopping experience

A further set of features and m-services is intended to improve the SQ for mobile shopping. They should make the mobile shopping for the customers more convenient and efficient. Although, most of these features can also be used inside a brick-and-mortar store, they are likely to improve the MASQ on the go.

F09: Landscape mode

Many customers complained that interacting with the three mobile shopping companion apps under investigation on a smartphone, it appears that they are not able to switch to landscape mode if the smartphone is turned into a horizontal position. Unlike most other mobile apps, the mobile shopping companion apps do not respond to the user moving the smartphone into another orientation which is measured by sensors. The landscape mode is often used to display additional information or larger graphics which is only hardly possible with the narrow display size in portrait mode. Further, the landscape mode seems to be more convenient to some customers when using the mobile app for longer periods of time. Thus, mobile shopping companion apps should support a landscape mode that is automatically used when the customer changes the orientation of the smartphone. "Can't believe the app is not able to go into landscape mode."

F10: Cross-device linkage

Following the idea of providing an omni-channel experience for their customers, retailers should not only implement mobile shopping companion apps for mobile devices but also synchronize the customer information and their shopping across multiple devices. This synchronization should also involve stationary devices such as desktop computers or laptops and wearable devices such as smart watches. The customer should have the opportunity to use his or her preferred device for making, continuing or finishing a purchase or information

[13] Mobile wallets store the user's credit, debit, prepaid, store and loyalty cards and other ticket information and make them available from the lock-screen without the need for starting the dedicated mobile app based on temporal and location-related information.

retrieval.

"The app has been updated to remember the content of my basket when changing from iPhone to iPad."

F11: Online purchase

Besides the features presented, a major functionality of a mobile shopping companion app is providing access to the retailer's online shop for mobile shopping. Although mobile devices have smaller displays and a different handling compared to desktop computers, customers should perceive a similar or even better shopping experience when shopping mobile using their mobile shopping companion app.

"I reserved tables and went to pick up, they charged me and said they would arrive in days when I can literally walk over and grab them in store(which I did) they could not refund my $ and I still have to come back and pickup then return the other tables."

F12: Favorites list and F13: Continuous basket

Another group of features deals with the virtual shopping basket of the mobile shopping companion app. The virtual shopping basket for storing the items already selected by the customer should remember the inserted items even the customer closes the app intentionally or unintentionally (e.g. when receiving a phone call) and in case of a crash of the overall mobile app. Besides, customers frequently request a shopping list feature. It should store the items the customer is going to purchase on his or her next shopping trip and should not only replace paper-based lists but extend it with additional features. For instance, it could calculate or estimate the total sum based on online and in-store item prices so that the customer knows in advance how much the purchase will cost. From this shopping list, items should also be markable as favorites to allow customers to easily add them for further shopping.

"It would also be great to have the option to save favourite items or remember previous searches."

F14: Item amendment

Analogous to the shopping list, items that the customer is willing of or thinking of buying in any future purchase should be stored on a so-called wish list. This list needs to be changeable if the customer decides to remove an item. Moreover, already placed orders should be changeable until the order is sent out to the customer. Changing an existing order means deleting or amending an item from the order or modifying the amount of a single item. Thus, the order should only be automatically finalized shortly before its delivery.

"This app is easy to use and very 'forgiving' - so if you make a mistake or change your mind about something it is easily amended - without having to start all over again."

F15: Swipe pictures

For items with more than one picture in the description, the customers demand to swipe through them without the necessity of closing the enlarged view of the item. A further improvement of the swiping feature would be the integration of 360 degree view enabling the customer to watch the product from every angle investigating each detail of the product.

"I would like the addition of being able to see a 360 view of products rather than several photos."

F16: Loyalty card manager

The mobile shopping companion app should not only encapsulate a digital representation of the loyalty card as described above, but also allow to administer the customer's whole loyalty program account on the move. This includes adding and changing personal information, checking the current status and automatically applying the points and discounts on purchases.

"It also Makes checking and adding 'Sparks' so easy. I've just shopped the sale early thanks to sparks."

6. Discussion

The present thesis follows a twofold approach consisting of a SLR for developing the MASQ measurement scale based upon previous literature and a QCA for deriving implementation guidelines and service design requirements for possible measures to improve the SQ of mobile shopping companion apps. The developed MASQ measurement scale follows a hierarchical and multi-dimensional approach based on recent literature (e.g., Stiakakis and Petridis, 2014; Vlachos et al., 2011). Hence, the MASQ of a mobile shopping companion app is measured more detailed using three primary dimensions, seven secondary dimensions and 22 items on the most detailed level. In the analyzed literature consensus seems to be achieved only regarding the primary dimensions interaction, environment and outcome quality (Fullerton, 2005; Lu et al., 2009; Stiakakis and Petridis, 2014). The secondary dimensions vary widely across the analyzed papers and thus are synthesized for the MASQ measurement scale. These are responsiveness, information, security and privacy, design, performance, technical reliability and valence (for the 22 items see ch. 4). Using the MASQ measurement scale as a framing, 22 implementation guidelines and 14 service design requirements are derived from the 10,099 analyzed online customer reviews of three exemplary mobile shopping companion apps in the QCA for the 22 measurement items. These implementation guidelines and service design requirements suggest areas to be improved to achieve a high MASQ. A high MASQ also requires a diverse set of features to be included in the mobile shopping companion app. At least 16 features and m-services from the three areas mobile app dissemination, seamless integration into brick-and-mortar environment and improving the mobile shopping experience should be integrated.

Implications for research

This thesis contributes to the extant literature by carrying forward the research on SQ to the context of MASQ. The MASQ measurement scale is tailored to assess the SQ of mobile shopping companion apps provided by retailers also operating brick-and-mortar stores. While previous papers measure ESQ, MSQ and MASQ using appropriate dimensions, some more recent papers introduce a hierarchical structure involving primary and secondary dimension like the developed MASQ measurement scale for mobile shopping com-

panion apps (e.g., Stiakakis and Petridis, 2014; Vlachos et al., 2011). The first investigations for a hierarchical structure for a SQ measurement scale originally stem from research by Dabholkar et al. (1996). It allows for a more detailed measurement of SQ based on the framing of generally accepted primary dimensions.

The developed MASQ measurement scale does not include changeable dimensions for different product categories. Huang et al. (2015) introduced different MSQ dimensions for virtual and tangible products. According to the authors, the privacy dimension is only required for measuring the MSQ for virtual products. However, including the protection and collection of the customers' personal data as SQ dimension is also essential for tangible products in the m-service environment because the same customer information have to be processed as for purchasing virtual products. When dealing with m-services on mobile devices data security is always a sensitive issue regardless of pure and non-pure services (Kuo et al., 2016, p. 5).

The extensive description of the dimensions and related items should encourage other researchers to apply the developed MASQ measurement scale on further mobile shopping companion apps and explicitly tests its relevance and validity on mobile shopping companion apps in an empirical manner. Further, it may be the basis for discussing the universality of the dimensions and items for mobile apps in the m-Commerce environment. The majority of analyzed papers close after the development of an ESQ, MSQ or MASQ measurement scale, without providing concrete suggestions for improving the quality of services. The implementation guidelines and service design requirements provided in this thesis extend the previous literature with concrete suggestions for improving the MASQ of mobile shopping companion apps in line with Knote and Söllner (2017) who provide these service design requirements for context-aware services in mobile navigation apps. Future research should test these suggestions for SQ improvements from an academic point of view and especially evaluate them with findings from other domains such as UI or user experience design.

As the online customer reviews exported from the App Store seem to be longer and more detailed than the ones crawled from the Play Store in this sample, it might be of future interest to further investigate the influence of a particular MADP on the quality of online customer reviews. Some customers reviewed that the SQ of the mobile shopping companion app is superior to the retailer's mobile website. Because this thesis focuses on the investigation of online customer reviews dealing with mobile shopping companion apps, future research may compare the user experience of mobile apps and mobile websites to examine whether it is worthwhile for retailers implementing dedicated mobile apps instead of offering a website responsive for mobile devices for their m-services provision.

Implications for practice

Practitioners can use the MASQ measurement scale to assess the SQ of their m-services and mobile shopping companion app and even benchmark their m-service with competitors. The dimensions and items can already give an indication on the areas considered as important for a high MASQ as expected by the customers. The derived implementation guidelines and service design requirements are especially interesting for practitioners. Retailers can try to optimize the quality of their m-services and their mobile shopping companion app according the identified 36 areas of optimization. Optimization is necessary at least for parts of the mobile apps as Rudolph et al. (2015) outline that mobile shopping companion apps bear potential for improvements concerning the implementation of new features and improvement of already existing m-services (p. 46). Providing a high MASQ mobile shopping companion app is essential for retailers fostering the in-store and online shopping at the same time. Today, customers invest 66 percent of the time they spend in m-Commerce activities on mobile devices but spend 78 percent of their money using desktop computers (Lipsman, 2017, p. 38). Converting at least some time into money spend on mobile devices will increase the revenue generated through m-Commerce. Additionally, the QCA reveals that customers tend to spend more money than intended if the mobile shopping companion app has a high MASQ and is especially easy to navigate ("This app is actually too easy... So easy to use it is going to cost me a lot of money?"). The proposed set of features can be a reference for retailers as to which features and m-services customers expect to be integrated in a mobile shopping companion app. Although each retailer should identify its required set of features individually, these suggested features and m-services can be a first indication (Rudolph et al., 2015, p. 43). Especially as customers prefer to have only one mobile shopping companion app instead of a variety of mobile apps dedicated to small sets of features.

Limitations

The present thesis involves some limitations and assumptions from a methodological and content point of view. For measuring the MASQ the disconfirmation paradigm is applied as already used by Grönroos (1984) and Parasuraman et al. (1985) for their initial researches on SQ. Indeed, SQ is observed as the difference between expected and perceived SQ by the customer. Since this time, there has been a discussion in literature on the acceptance and validity of this general paradigm as basis for measuring SQ and the applicability of single dimensions of the measurement scales (e.g., Buttle, 1996; Cronin and Taylor, 1994). As the MASQ measurement scale also applies the disconfirmation paradigm, it is also open to the criticism on the evidence for the definition of SQ as the gap between perceived and expected SQ and the focus on the service delivery only rarely takes into account the outcome of the service encounter (Buttle, 1996, p. 10). Nevertheless, the disconfirmation paradigm is also applied in the majority of scales measuring ESQ and MSQ. Thus, it is considered to be also valid in the context of MASQ.

The MASQ measurement scale is based upon extant literature on ESQ and MSQ identified in a SLR. Thus, the validity of the existing dimensions is mostly empirically proven

in previous literature. However, the transferability of the primary and secondary dimensions and the items taken from the extant literature to the context of mobile shopping companion apps is not explicitly empirically proven. Nevertheless, their relevance and validity in the context of MASQ is implicitly shown with the identification of relevant online customer reviews and the derivation of implementation guidelines and service design requirements for each of the items. Analogous, the relevance and validity is implicitly shown for secondary dimensions and items that are not taken from previous literature but derived in the QCA. Moreover, the MASQ measurement scale does not include a questionnaire or thesis paper that can be directly applied by the retailers to assess the SQ of their mobile shopping companion apps. Instead, for measuring the MASQ the extensive dimension and item descriptions have to be utilized without indicating unambiguous scale values.

To derive implementation guidelines and service design requirements, online customer reviews are analyzed in a QCA introduced by (Mayring, 2015). Using online customer reviews, a form of eWOM, as the basis for the analysis involves some limitations. Firstly, customers usually do not spend much time to provide systematic feedback but rather want to give their opinion quickly (Pagano and Maalej, 2013, p. 130). On the one hand, this feedback tends to be honest and straight, but on the other hand it can be confusing with a furious overtone. Secondly, among the authentic online customer reviews published by real customers and users of the mobile shopping companion app, there can be spam and misleading reviews (Genc-Nayebi and Abran, 2017, p. 1). Further, authors of online customer reviews act in anonymity publishing their reviews using nicknames (Trenz and Berger, 2013, p. 2). Therefore, also developers or providers of mobile apps can write online customer reviews anonymously to increase the awareness of the mobile apps in MADPs or to improve the rating of the mobile apps. This phenomenon is called "review manipulation or fraud" (Trenz and Berger, 2013, p. 3). If these fraud reviews are not identified correctly and excluded during the QCA, they are able to distract the results of the analysis. Thirdly, for complaints about problems and glitches it cannot be differentiated whether it is an issue concerning the mobile app, an issue with the customer's mobile device executing the mobile app, an intertwined third-party m-service or the customer's ability to operate the mobile app correctly. Fourthly, the 10,099 exported online customer reviews cannot be considered as a representative sample. They are just the most recent online customer reviews available for the mobile shopping companion apps implemented by Walmart, Tesco and M&S available at the MADPs App Store and Play Store. For delivering useful results, these online customer reviews have to be current because the SQ and features of the mobile apps can change rapidly with frequent releases.

Moreover, the actual methodological application of the QCA by Mayring (2015) in this thesis is based on certain assumptions and causes further limitations. During the QCA the online customer reviews are not linked to a specific version of the mobile app anymore although the version number is provided in the initial exports from the MADPs. Thus, it is assumed for the QCA that all the online customer reviews of a specific mobile shopping companion app are related to a single release of the mobile app. Taking into account the version number in the QCA would have been helpful because the mobile shopping companion apps are updated frequently with a varying SQ and set of features (Nayebi et al., 2016). Hence, each online customer review only applies to a specific release of the mobile app. Oftentimes reviewers point out the improvements compared to the previous releases of the mobile app. One user claims for instance with reference to previous releases in mind that the recent version of the M&S app "seems to be best yet". Fu et al. (2013) even proposes to specify the version number of the mobile app when analyzing mobile apps and related online customer reviews.

Although online customer reviews can be rated and reviewed in the MADPs as kind of a "quality criteria" (Mayring, 2015, p. 53), reviews rating and reviewing the analyzed online customer reviews are not taken into account for the QCA. Initially, the online customer reviews are treated as equally relevant for the QCA. These quality criteria might be an indication for the relevance of an online customer review for other customers. Hence, in the applied QCA approach online customer reviews not relevant from the perspective of other customers might influence the result of the QCA. Citing special forms or representative examples from the set of online customer reviews may involve the danger of generalizations or distortions suggesting uniformity and polarity that do not exist in the initial sample (Mayring, 2015, p. 106). Thus, citing from exemplary online customer reviews from the prototyping approach identified during the QCA might provide a wrong impression of the overall sample of online customer reviews exported from the MADPs. For this thesis, a manual process for analyzing the online customer reviews is undertaken. This process is not efficient enough for practitioners for analyzing the steadily increasing amount of ratings and online customers reviews published in the MADPs (Pagano and Maalej, 2013, p. 134). Hence, it might be of interest for providers of mobile apps to develop a fully automated process for the QCA from the crawling of online customer reviews from the MADPs to the summary of the results per category using advanced text mining tools. Applying an automated process would result in reproducible results that are not biased by personal opinions and subjective evaluations of the people executing the manual analysis. However, it has to cope with the difficulty of filtering ironic comments from the online customer reviews (Pagano and Maalej, 2013, p. 130). As the focus of the thesis lies on the implementation guidelines and service design requirements resulting from the QCA and not on the implementation of an automatic analysis process involving text mining techniques, a manual approach is followed for analyzing and categorizing the 10,099 online customer reviews from the App Store and Play Store (sec. 3.2).

Although this thesis involves mobile shopping companion apps implemented by retailers from two different countries

and culture might also have an influence on the perceived MASQ, this thesis does not differentiate among different cultures, their influences on the MASQ and its impact on the implementation guidelines and service design requirements for mobile shopping companion apps. Tsikriktsis (2002) revealed that culture has an influence on the perceived ESQ and that especially the dimensions masculinity and long-term orientation identified by Hofstede (1991) expect a higher WDQ.

7. Conclusion and Outlook

Similar to the e-service environment, SQ is an important determinant in the context of m-service in the m-Commerce environment for customer loyalty, revenue generated with these services and overall company success. This also applies for mobile shopping companion apps disseminated by retailers that also operate brick-and-mortar stores besides their electronic and mobile businesses. Mobile shopping companion apps extend the traditional and e-services of these retailers by an additional mobile channel supporting the shopping within and outside the brick-and-mortar stores using mobile technology. Overall, these mobile apps are designated to enhance the customers' shopping experience.

Reviewing the extant literature on the SQ of mobile apps in the m-Commerce context, there is only little research published dealing with the measurement of MASQ. Thus, a scale for measuring the MASQ has to draw on previous findings in ESQ and MSQ and existing dimensions need even to be adapted to the context of mobile shopping companion apps. The measurement of the MASQ is based on the customers' expectations and perceptions of the SQ offered by the mobile shopping companion app following the disconfirmation paradigm. Moreover, in the literature on ESQ, MSQ and MASQ neither a consensus is achieved on the number and nature of dimensions used for measuring the SQ nor on the general denotations of the dimensions or on their definitions. However, all the identified SQ measurement scales apply a multi-dimensional approach. Only more recent measurement scales applying a hierarchical approach commonly share the three primary dimensions interaction quality, environment quality and outcome quality. Although these scales have the framing of the primary dimension in common, the secondary dimensions and related items vary. Hence, the measurement of the SQ of mobile shopping companion apps requires a scale tailored to the m-Commerce environment and the mobile shopping companion app supporting the customer's online shopping and also enhancing the customer's in-store shopping experience. Additionally, the measurement scale needs to cover the whole shopping process from the information to the after-sales phase. Hence, the MASQ measurement scale applying a hierarchical and multi-dimensional approach for measuring the SQ of mobile shopping companion apps is developed. It consists of three primary dimensions, seven secondary dimensions and 22 related items on the most detailed level. Using this scale, the MASQ is measured by assessing the SQ of a mobile shopping companion app using the description of the items and contracting the results upwards the hierarchy.

As the MASQ measurement scale is based on the disconfirmation paradigm, the cus- tomers' expectations and perception about the offered SQ have to be taken into account. Therefore, online customer reviews of the mobile shopping companion apps by Walmart, Tesco and M&S as a form of eWOM are exported from two MADPs. These online customer reviews encapsulating the customers' expectations and perceptions of the MASQ and thus are analyzed to derive design requirements for achieving a high MASQ for mobile shopping companion apps. The design requirements are further subdivided into implementation guidelines affecting the actual design and implementation of the mobile shopping companion app and the service design requirements dealing with specification of the m-services provided through the mobile shopping companion app. Deduced from the QCA , 22 implementation guidelines and 14 service design requirements are identified as possible areas for optimizing the MASQ and achieving a high MASQ when developing mobile shopping companion apps. The areas of optimization affect the design, development, implementation, maintenance and operation of the mobile app. The QCA follows a content structuring approach as a form of a deductive category assignment. The QCA further reveals that although each retailer should decide specifically on the set of features and m-services provided in the mobile shopping companion app, a set of 16 exemplary features and m-services should be included in an arbitrary mobile shopping companion app. These features and m-services should be encapsulated in a single mobile shopping companion app and not spread across multiple m-Commerce-related mobile apps.

Future research should scrutinize and strive to validate the results of this thesis in a broader context. Firstly, as the developed MASQ measurement scale encapsulates SQ dimensions previously introduced in existing and renowned measurement scales for ESQ and MSQ, future research should explicitly and empirically test the applicability and validity of the proposed primary and secondary dimensions and items in the context of mobile shopping companion apps by measuring and comparing the MASQ of several mobile shopping companion apps. Secondly, it should be assessed if the proposed dimensions and items measure the MASQ of an arbitrary mobile shopping companion app entirely. Thirdly, the potential applicability of the MASQ dimensions and items on more general mobile shopping apps should be tested. This investigation should result in proposals for adaptions and extensions possibly required to make it applicable on general mobile shopping apps. Possible further dimensions explicitly excluded from the MASQ measurement scale deal with the quality of the delivery and the product. Fourthly, future research should extend the developed MASQ measurement scale with a questionnaire including questions and statements for directly and unambiguously assessing the SQ within each dimension and item. This summarized MASQ measurement scale should simplify the MASQ measurement for practitioners. Fifthly, the impact of the suggested imple-

mentation guidelines and service design requirements on the MASQ of mobile shopping companion apps should be tested. This test should involve the perceptions of a sample of customers on the suggested adaptions of the mobile shopping companion app and m-services. In addition, the proposed set of features should be reviewed taking into account the set of features of further mobile shopping companion apps and features and m-services requested by the customers of these further mobile apps.

References

148apps.com. App Detail: Walmart - Shopping and Saving, 2017. URL http://www.148apps.com/app/338137227/.

Aladwani, A. M. Developing and Validating an Instrument for Measuring User-Perceived Web Quality. *Information & Management*, 39(6):467–476, 2002.

Alpar, P., Alt, R., Bensberg, F., Grob, H. L., Weimann, P., and Winter, R. *Anwendungsorientierte Wirtschaftsinformatik: Strategische Planung, Entwicklung und Nutzung von Informationssystemen*. 6. Edition, Wiesbaden, Vieweg + Teubner, 2011.

Balauca, C. and Mehta, G. Exploring the Potential of Social Media Tools to Enhance Customer's Satisfaction and Increase Turnover in the Irish Accommodation Sector. In *Proceedings of the11th Annual Tourism and Hospitality Research in Ireland Conference (THRIC '15)*, Letterkenny, Ireland, 2015.

Barnes, S. J. and Vidgen, R. T. An Integrative Approach to the Assessment of e-Commerce Quality. *Journal of Electronic Commerce Research*, 3(2): 114–127, 2002.

Bauer, H. H., Falk, T., and Hammerschmidt, M. eTransQual: A Transaction Process-Based Approach for Capturing Service Quality in Online Shopping. *Journal of Business Research*, 59(7):866–875, 2006.

Bitner, M. J. Evaluating Service Encounters: The Effects of Physical Surroundings and Employee Responses. *Journal of Marketing*, 54(2):69–82, 1990.

Blut, M. E-Service Quality: Development of a Hierarchical Model. *Journal of Retailing*, 92(4):500–517, 2016. ISSN 00224359. doi: 10.1016/j.jretai.2016.09.002. URL http://dx.doi.org/10.1016/j.jretai.2016.09.002.

Blut, M., Chowdhry, N., Mittal, V., and Brock, C. E-Service Quality: A Meta-Analytic Review. *Journal of Retailing*, 91(4):679–700, 2015.

Bouwman, H., Bejar, A., and Nikou, S. Mobile Services Put in Context: A Q-Sort Analysis. *Telematics and Informatics*, 29(1):66–81, 2012.

Brady, M. K. and Cronin, J. J. Perceived Service Conceptualizing Approach Quality: A Hierarchical Approach. *Journal of Marketing*, 65(3):34–49, 2001.

Bressolles, G., Durrieu, F., and Giraud, M. The Impact of Electronic Service Quality's Dimensions on Customer Satisfaction and Buying Impulse. *Journal of Customer Behaviour*, 6(1):37–56, 2007.

Brown, M. Tesco Embraces iPhone with 'Groceries' App, 2010. URL http://www.wired.co.uk/article/tesco-app.

Bundesnetzagentur. Entwicklung des Datenvolumens im Mobilfunk in Deutschland von 2005 bis 2016 (in Millionen Gigabyte), 2017. URL https://de.statista.com/statistik/daten/studie/172798/umfrage/datenvolumen-im-deutschen-mobilfunkmarkt-seit-2005/.

Buttle, F. SERVQUAL: Review, Critique, Research agenda. *European Journal of Marketing*, 30(1):8–32, 1996.

Cai, S. and Jun, M. Internet Users' Perceptions of Online Service Quality: A Comparison of Online Buyers and Information Searchers. *Managing Service Quality: An International Journal*, 13(6):504–519, 2003.

Cao, M., Zhang, Q., and Seydel, J. B2C e-Commerce Web Site Quality: An Empirical Examination. *Industrial Management & Data Systems*, 105(5): 645–661, 2005.

Charland, A. and Leroux, B. Mobile Application Development: Web vs . Native. *Communications of the ACM*, 54(5):49–53, 2011.

Clark, B. Welcome to My Parlor: The Lure of Marketing on the World Wide Web is Great. Be Sure You Don't Get Stuck with the Wrong Approach. *Marketing Management*, 5(4):11–22, 1997.

Clarke, I. Emerging Value Propositions for m-Commerce. *Journal of Business Strategies*, 18(4):464–494, 2001.

Clarke, I. and Flaherty, T. Mobile Portals: The Development of m-Commerce Gateways. In Mennecke, B. and Strader, T., editors, *Mobile Commerce: Technology, Theory, and Applications*. Idea Group Publishing, Hershey: 185-201, 2003.

Collier, J. E. and Bienstock, C. C. Measuring Service Quality in E-Retailing. *Journal of Service Research*, 8(3):260–275, 2006.

Cooper, H. M. Organizing Knowledge Synthesis: A Taxonomy of Literature Reviews. *Knowledge in Society*, 1(1):104–126, 1988.

Cox, J. and Dale, B. G. Service Quality and e-Commerce: An Exploratory Analysis. *Managing Service Quality: An International Journal*, 11(2):121–131, 2001.

Cristobal, E., Flavián, C., and Guinalíu, M. Perceived e-Service Quality (PeSQ): Measurement Validation and Effects on Consumer Satisfaction and Web Site Loyalty. *Managing Service Quality: An International Journal*, 17(3):317–340, 2007.

Cronin, J. J. and Taylor, S. A. SERVPERF versus SERVQUAL: Reconciling Performance-Based and Perceptions-Minus-Expectations Measurement of Service Quality. *Journal of Marketing*, 58(1):125–131, 1994.

Dabholkar, P. A., Thorpe, D. I., and Rentz, J. O. A Measure of Service Quality for Retail Stores: Scale Development and Validation. *Academy of Marketing Science*, 24(1):3–16, 1996.

Dibia, V. and Wagner, C. Success within App Distribution Platforms: The Contribution of App Diversity and App Cohesivity. In *Proceedings of the 48th Annual Hawaii International Conference on System Sciences (HICSS '15)*, pages 4304–4313, Kauai, Hawaii, USA, 2015. ISBN 9781479973675. doi: 10.1109/HICSS.2015.515.

Ding, D. X., Hu, P. J. H., and Sheng, O. R. L. E-SELFQUAL: A Scale for Measuring Online Self-Service Quality. *Journal of Business Research*, 64 (5):508–515, 2011.

Eaton, J. and Bawden, D. What Kind of Resource is Information? *International Journal of Information Management*, 11(2):156–165, 1991.

EMarketer. Mcommerce's Rapid Growth Is Primarily Coming from Smartphones, 2016. URL https://www.emarketer.com/Article/Mcommerces-Rapid-Growth-Primarily/Coming-Smartphones/1013909.

EMarketer. Umsatz im Mobile Commerce in den USA in den Jahren 2011 bis 2015 sowie eine Prognose bis 2020 (in Milliarden US-Dollar), 2017. URL https://de.statista.com/statistik/daten/studie/208928/umfrage/umsatz-im-mobile-commerce-in-den-usa-und-prognose/.

Evans, J. R. and Lindsay, W. M. *Managing for Quality and Performance Excellence*. South-Western, Mason, 9. edition, 2011.

Fassnacht, M. and Koese, I. Quality of Electronic Services. *Journal of Service Research*, 9(1):19–37, 2006.

Faulkner, R. Marks & Spencer Launches iPhone App, 2012. URL https://www.drapersonline.com/marks-and-spencer-launches-iphone-app/5038606.article.

Fischer, P. and Hofer, P. *Lexikon der Informatik*. 14. Edition, Berlin Heidelberg, Springer-Verlag, 2008.

Francis, J. E. Category-specific RECIPEs for Internet Retailing Quality. *Journal of Services Marketing*, 23(7):450–461, 2009.

Francis, J. E. and Lesley, W. Exploratory and Confirmatory Factor Analysis of the Perceived Internet Retailing Quality Model. In *Proceedings of the 17th Australian and New Zealand Marketing Academy Conference (ANZMAC '02): 1791-1796*, Melbourne, 2002.

Fu, B., Lin, J., Li, L., Faloutsos, C., Hong, J., and Sadeh, N. Why People Hate Your App: Making Sense of User Feedback in a Mobile App Store. In *Proceedings of the 19th International Conference on Knowledge Discovery and Data Mining (SIGKDD '13)*, pages 1276–1284, Chicago, USA, 2013. ISBN 9781450321747. doi: 10.1145/2487575.2488202. URL http://dl.acm.org/citation.cfm?id=2488202.

Fullerton, G. The Service Quality-Loyalty Relationship in Retail Services: Does Commitment Matter? *Journal of Retailing and Consumer Services*, 12(2):99–111, 2005.

Gefen, D. Customer Loyalty in e-Commerce. *Journal of the Association for Information Systems*, 3(1):27–51, 2002.

Genc-Nayebi, N. and Abran, A. A Systematic Literature Review: Opinion Mining Studies from Mobile App Store User Reviews. *Journal of Systems and Software*, 125:207–219, 2017.

Giannakos, M. N. Absolute Price as a Determinant of Perceived Service Quality in Hotels: A Qualitative Analysis of Online Customer Reviews. *International Journal of Hospitality and Event Management*, 1(1):62–80, 2014.

Grewal, D., Roggeveen, A. L., and Nordfällt, J. The Future of Retailing. *Journal of Retailing*, 93(1):1–6, 2017.

Grönroos, C. A Service Quality Model and its Marketing Implications. *European Journal of Marketing*, 18(4):36–44, 1984.

Groß, M. Mobile shopping: A Classification Framework and Literature Review. *International Journal of Retail & Distribution Management*, 43(3): 221–241, 2015.

Gummerus, J. and Pihlström, M. Context and Mobile Services Value-in-Use. *Journal of Retailing and Consumer Services*, 18(6):521–533, 2011.

Heinonen, K. and Pura, M. Classifying Mobile Services. *All Sprouts Content*, 6(42):1–16, 2008.

Heinrich, L. J., Heinzl, A., and Roithmayr, F. *Wirtschaftsinformatik Lexikon*.

7. Edition, München, Oldenbourg Wirtschaftsverlag, 2004.

Hennig-Thurau, T., Gwinner, K. P., Walsh, G., and Gremler, D. D. Electronic Word-of-Mouth via Consumer-Opinion Platforms: What Motivates Consumers to Articulate Themselves on the Internet? *Journal of Interactive Marketing*, 18(1):38–52, 2004.

Hofstede, G. *Cultures and Organizations: Software of the Mind*. Maidenhead, McGraw-Hill, 1991. URL http://www.publidisa.com/PREVIEW-LIBRO-9780071439596.pdf.

Holloway, B. B. and Beatty, S. E. Satisfiers and Dissatisfiers in the Online Environment. *Journal of Service Research*, 10(4):347–364, 2008.

Hsu, C. *Service Science: Design for Scaling and Transformation*. World Scientific Publishing, Singapore, 2009.

Hsu, T., Hung, L., and Tang, J. The Multiple Criteria and Sub-Criteria for Electronic Service Quality Evaluation. *Online Information Review*, 36(2): 241–260, 2012.

Hu, Y. An Instrument for Measuring Service Quality in the e-Commerce Environment. In *Proceedings of the 1st International Conference on Innovative Computing and Communication (CICC '10) and Asia-Pacific Conference on Information Technology and Ocean Engineering (ITOE '10): 300-303*, 2010.

Huang, E. Y., Lin, S.-W., and Fan, Y.-C. M-S-QUAL: Mobile Service Quality Measurement. *Electronic Commerce Research and Applications*, 14(2): 126–142, 2015.

Idu, A., Zande, T. V. D., and Jansen, S. Multi-homing in the Apple Ecosystem: Why and How Developers Target Multiple Apple App Stores. In *Proceedings of the 3rd International Conference on Management of Emergent Digital EcoSystems (MEDES '11): 122-128*, San Francisco, USA, 2011.

International Organization for Standardization. *ISO 9000:2008. Management Systems - Fundamentals and Vocabulary*. ISO, Gevena, 2008.

Janda, S., Trocchia, P. J., Gwinner, K. P., and Authors, F. Consumer Perceptions of Internet Retail Service Quality. *International Journal of Service Industry Management*, 13(5):412–431, 2002.

Jun, M., Yang, Z., and Kim, D. Customers' Perceptions of Online Retailing Service Quality and Their Satisfaction. *International Journal of Quality & Reliability Management*, 21(8):817–840, 2004.

Kar, E. V. D., Maitland, C., de Montalvo, U. W., and Bouwman, H. Design Guidelines for Mobile Information and Entertainment Services: Based on the Radio538 Ringtunes i-Mode Service Case Study. In *Proceedings of the 5th International Conference on Electronic Commerce (ICEC '05): 413-421*, Pittsburgh, USA, 2003.

Kettinger, W. J. and Lee, C. C. Pragmatic Perspectives on the Measurement of Information Systems Service Quality. *MIS Quarterly*, 21(2):223–240, 1997.

Kettinger, W. J. and Lee, C. C. Zones of Tolerance: Alternative Scales for Measuring Information Systems Service Quality. *MIS Quarterly*, 29(4): 607–623, 2005.

Knote, R. and Söllner, M. Towards Design Excellence for Context-Aware Services - The Case of Mobile Navigation Apps. In *Proceedings of the 13th Internationale Tagung Wirtschaftsinformatik (WI '17): 121-135*, St. Gallen, Switzerland, 2017.

Kumar, R. A Proposed Scale of Assessing Mobile App Service Quality (MAPP-SQL). *ELK Asia Pacific Journal of Marketing and Retail Management*, 8(1): 1–11, 2017.

Kuo, T., Tsai, G. Y., Lu, I.-Y., and Chang, J.-S. Relationships among Service Quality, Customer Satisfaction and Customer Loyalty: A Case Study on Mobile Shopping APPs. In *Proceedings of the 17th Asia Pacific Industrial Engineering and Management Systems Conference (APIEMS '16): 1-8*, Taipei, Taiwan, 2016.

Ladhari, R. Developing e-Service Quality Scales: A Literature Review. *Journal of Retailing and Consumer Services*, 17(6):464–477, 2010.

Laudon, K. C., Laudon, J. P., and Schoder, D. *Wirtschaftsinformatik: Eine Einführung*. 2. Edition, München, Pearson Studium, 2010.

Lee, G. and Lin, H. Customer Perceptions of e-Service Quality in Online Shopping. *International Journal of Retail & Distribution Management*, 33 (2):161–176, 2005.

Lee, Y. E. and Benbasat, I. A Framework for the Study of Customer Interface Design for Mobile Commerce. *International Journal of Electronic Commerce*, 8(3):79–102, 2004.

Li, F. and Leng, Z. Study on the Quality of Electronic Services: Conceptualizing and Exploratory Analysis. In *Proceedings of the 4th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM '08)*, pages 1–4, Dalian, China, 2008.

Lin, Y., Luo, J., Cai, S., Ma, S., and Rong, K. Exploring the Service Quality in the e-Commerce Context: A Triadic View. *Industrial Management & Data Systems*, 116(3):388–415, 2016.

Lipsman, A. State of the U.S. Online Retail Economy in Q1 2017, 2017. URL http://www.comscore.com/Insights/Presentations-and-Whitepapers/2017/State-of-the-US-Online-Retail-Economy-in-Q1-2017.

Loiacono, E., Watson, R., and Goodhue, D. WebQual: An Instrument for Consumer Evaluation of Web Sites. *International Journal of Electronic Commerce*, 11(3):51–87, 2007.

Long, M. and McMellon, C. Exploring the Determinants of Retail Service Quality on the Internet. *Journal of Services Marketing*, 18(1):78–90, 2004.

Lu, Y., Zhang, L., and Wang, B. A Multidimensional and Hierarchical Model of Mobile Service Quality. *Electronic Commerce Research and Applications*, 8(5):228–240, 2009.

Lynch, J. G. and Ariely, D. Wine Online: Search Costs Affect Competition on Price, Quality, and Distribution. *Marketing Science*, 19(1):83–103, 2000.

Madu, C. N. and Madu, A. A. Dimensions of e-Quality. *International Journal of Quality & Reliability Management*, 19(3):246–258, 2002.

Mayring, P. *Qualitative Content Analysis: Theoretical Foundation, Basic Procedures and Software Solution*. Klagenfurt, Beltz, 2014.

Mayring, P. *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. 12. Edition, Weinheim, Beltz, 2015.

Meola, A. The Rise of m-Commerce: Mobile Shopping Stats & Trends, 2016. URL http://www.businessinsider.de/mobile-commerce-shopping-trends-stats-2016-10?r=US{&}IR=T.

Microsoft. Windows App Store, 2017. URL https://www.microsoft.com/en-us/store/apps/windows?icid=CNavAppsWindowsApps.

M&S. M&S 2017 Annual Report, 2017a. URL http://annualreport.marksandspencer.com/M{&}S_AR2017.pdf.

M&S. Key Facts, 2017b. URL http://corporate.marksandspencer.com/aboutus/key-facts.

M&S. M&S, 2017c. URL https://itunes.apple.com/gb/app/m-s/id538410698?mt=8.

Mudambi, S. M. and Schuff, D. What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com. *MIS Quarterly*, 34(1):185–200, 2010.

Nayebi, M., Adams, B., and Ruhe, G. Release Practices for Mobile Apps - What Do Users and Developers Think? In *Proceedings of the 23rd International Conference on Software Analysis, Evolution, and Reengineering (SANER '16), Volume 1: 552-562*, Osaka, Japan, 2016. doi: 10.1109/SANER.2016.116.

Nickerson, R., Muntermann, J., Varshney, U., and Isaac, H. Taxonomy Development in Information Systems: Developing a Taxonomy of Mobile Applications. In *Proceedings of the 17th European Conference on Information Systems (ECIS '09)*, pages 1–12, Verona, Italy, 2009.

Norman, R. *Service Management Strategy and Leadership*. New York, Wiley, 1984.

Pagano, D. and Maalej, W. User Feedback in the AppStore: An Empirical Study. In *Proceedings of the 21st International Requirements Engineering Conference (RE '13): 125-134*, Rio de Janeiro, Brazil, 2013.

Parasuraman, A., Zeithaml, V. A., and Berry, L. L. A Conceptual Model of Service Quality and Its Implications for Future Research. *The Journal of Marketing*, 49(4):41–50, 1985.

Parasuraman, A., Zeithaml, V. A., and Berry, L. L. A Multiple Item Scale for Measuring Consumer Perceptions of Service Quality. *Journal of Retailing*, 64(1):12–37, 1988.

Parasuraman, A., Zeithaml, V. A., and Berry, L. L. Reassessment of Expectations as a Comparison Standard in Measuring Service Quality: Implications for Further Research. *Journal of Marketing*, 58(1):111–124, 1994.

Parasuraman, A., Zeithaml, V. A., and Malhotra, A. E-S-Qual: A Multiple-Item Scale for Assessing Electronic Service Quality. *Journal of Service Research*, 7(3):213–233, 2005.

Perez, S. Walmart Goes Mobile with New Apps for iPhone & iPad, 2011. URL https://techcrunch.com/2011/11/08/walmart-goes-mobile-with-new-apps/.

Pitt, L. F., Watson, R. T., and Kavan, C. B. Service Quality: A Measure of Information System Effectiveness. *MIS Quarterly*, 19(2):173–187, 1995.

Pohl, K. *Requirements Engineering: Fundamentals, Principles, and Techniques*.

Berlin, Heidelberg, Springer-Verlag, 2010.

Put, A., Dacosta, I., Milutinovic, M., Decker, B. D., Seys, S., Boukayoua, F., Naessens, V., Vanhecke, K., Pessemier, T. D., and Martens, L. inShopnito: An Advanced yet Privacy-Friendly Mobile Shopping Application. *2014 IEEE World Congress on Services*, pages 129–136, 2014.

Rao, S. and Troshani, I. A Conceptual Framework and Propositions for the Acceptance of Mobile Services. *Journal of Theoretical and Applied Electronic Commerce Research*, 2(2):61–73, 2007.

Ravindranath, L., Padhye, J., Agarwal, S., Mahajan, R., and Al, E. AppInsight: Mobile App Performance Monitoring in the Wild. In *Proceedings of the 10th USENIX Conference on Operating Systems Design and Implementation (OSDI' 12): 107-120*, Hollywood, USA, 2012.

Reeves, C. A. and Bednar, D. A. Defining Quality: Alternatives and Implications. *The Academy of Management Review*, 19(3):419–445, 1994.

Riedl, C., Leimeister, J. M., and Krcmar, H. New Service Development for Electronic Services: A Literature Review. In *Proceedings of the 15th Americas Conference on Information Systems (AMCIS '09): 1-9*, San Francisco, USA, 2009.

Rigby, D. The Future of Shopping. *Harvard Business Review*, 89(12):65–76, 2011.

Rosen, D. E. and Purinton, E. Website Design: Viewing the Web as a Cognitive Landscape. *Journal of Business Research*, 57(7):787–794, 2004.

Rudolph, T., Nagengast, L., Melanie, B., and Bouteiller, D. Die Nutzung mobiler Shopping Apps im Kaufprozess. *Marketing Review St. Gallen*, 3 (32):42–49, 2015.

Rust, R. T. and Oliver, R. L. Service Quality: Insights and Managerial Implications from the Frontier. In Rust, R. T. and Oliver, R. L., editors, *Service Quality: New Directions in Theory and Practice*, pages 1–19. Sage Publications, Thousand Oaks, 1994. URL https://books.google.de/books?hl=en{&}lr={&}id=c3woDAAAQBAJ{&}oi=fnd{&}pg=PT7{&}dq=Service+Quality:+New+Directions+in+Theory+and+Practice{&}ots=zo3fyfqOGu{&}sig=Qaqq1088yK4ZLtlAMTrblbFHXQg{#}v=onepage{&}q=ServiceQuality{%}3ANewDirectionsinTheoryandPractice{&}f=false.

Rust, R. T. and Lemon, K. N. E-Service and the Consumer e-Service and the Consumer. *Journal of Electronic Commerce*, 5(3):85–101, 2001.

Santos, J. E-Service Quality: A Model of Virtual Service Quality Dimensions. *Managing Service Quality: An International Journal*, 13(3):233–246, 2003.

Seth, A., Gupta, H. M., and Momaya, K. Quality of Service Parameters in Cellular Mobile Communication. *International Journal of Mobile Communications*, 5(1):68–93, 2007.

Seth, N., Deshmukh, S., and Vrat, P. Service Quality Models: A Review. *International Journal of Quality & Reliability Management*, 22(9):913–949, 2005.

Sheth, J. N., Newman, B. I., and Gross, B. L. *Consumption Values and Market Choices: Theory and Applications*. Cincinnati, South-Western, 1991.

Silberer, G. *Mobile Commerce: Grundlagen, Geschäftsmodelle, Erfolgsfaktoren*. Wiesbaden, Gabler, 2002.

Sohn, C. and Tadisina, S. K. Development of e-Service Quality Measure for Internet-Based Financial Institutions. *Total Quality Management & Business Excellence*, 19(9):903–918, 2008.

Solomon, B. Shopping Apps Are Now the Fastest Growing Thing in Mobile, 2015. URL https://www.forbes.com/sites/briansolomon/2015/01/06/shopping-apps-are-now-the-fastest-growing-thing-in-mobile/{#}4b39f412f738.

Sommerville, I. *Software Engineering*. 9. Edition, Boston, Addison-Wesley, 2011.

Sower, V., Duffy, J., Kilbourne, W., Kohers, G., and Jones, P. The Dimensions of Service Quality For Hospitals: Development and Use of the KQCAH Scale. *Health Care Management Review*, 26(2):47–59, 2001.

Spreer, P. and Rauschnabel, P. A. Selling with Technology: Understanding the Resistance to Mobile Sales Assistant Use in Retailing. *Journal of Personal Selling & Sales Management*, 36(3):240–263, 2016.

Stahlknecht, P. and Hasenkamp, U. *Einführung in die Wirtschaftsinformatik*. 11. Edition, Berlin, Springer, 2005.

Stiakakis, E. and Petridis, K. Developing and Validating a Multi-Criteria Model to Evaluate Mobile Service Quality. In Martínez-López, F., editor, *Handbook of Strategic e-Business Management*, pages 935–956. Springer-Verlag, Berlin Heidelberg, 2014. doi: 10.1007/978-3-642-39747-9. URL http://link.springer.com/10.1007/978-3-642-39747-9.

Stiakakis, E., Georgiadis, C., and Petridis, K. Analyzing m-Service Quality Dimensions Using Multivariate Statistical Techniques. In *Proceedings of the 12th International Conference on Mobile Business (ICMB '13)*, number 24, Berlin, Germany, 2013. URL http://aisel.aisnet.org/icmb2013/24.

Swaid, S. I. and Wigand, R. T. Measuring the Quality of e-Service: Scale Development and Initial Validation. *Journal of Electronic Commerce Research*, 10(1):13–28, 2009.

Tan, F. B. and Chou, J. P. C. The Relationship between Mobile Service Quality, Perceived Technology Compatibility, and Users' Perceived Playfulness in the Context of Mobile Information and Entertainment Services. *International Journal of Human-Computer Interaction*, 24(7):649–671, 2008.

Taylor, D. G. and Levin, M. Predicting Mobile App Usage for Purchasing and Information-Sharing. *International Journal of Retail & Distribution Management*, 42(8):759–774, 2014.

Tesco. Tesco 2017 Anual Report, 2017a. URL https://www.tescoplc.com/media/392373/68336_tesco_ar_digital_interactive_250417.pdf.

Tesco. Contacts, 2017b. URL https://www.tescoplc.com/contacts/general/.

Tesco. History, 2017c. URL https://www.tescoplc.com/about-us/history/.

Trenz, M. and Berger, B. Analyzing Online Customer Reviews - An Interdisciplinary Literature Review And Research Agenda. In *Proceedings of the 21st European Conference on Information Systems (ECIS '13)*, pages 1–12, Utrecht, Netherlands, 2013. doi: 10.13140/2.1.1203.1361. URL http://aisel.aisnet.org/ecis2013{_}cr/83{%}5Cnhttp://aisel.aisnet.org/cgi/viewcontent.cgi?article=1306{&}context=ecis2013{_}cr{%}5Cnhttp://aisel.aisnet.org/ecis2013{_}cr/83/.

Tsikriktsis, N. Does Culture Influence Website Quality Expectations? An Empirical Study. *Journal of Service Research*, 5(2):1–39, 2002.

van Dyke, T. P., Kappelman, L. A., and Prybutok, V. R. Measuring Information Systems Service Quality: Concerns on the Use of the SERVQUAL Questionnaire. *MIS Quarterly*, 21(2):195–208, 1997.

Varshney, U. and Vetter, R. Mobile Commerce: Framework, Applications and Networking Support. *Mobile Networks and Applications*, 7:185–198, 2002.

VATM. Durchschnittliches Datenvolumen im Monat pro Mobilfunkanschluss* in Deutschland von 2009 bis 2016 (in Megabyte), 2017. URL https://de.statista.com/statistik/daten/studie/3506/umfrage/monatliches-datenvolumen-pro-mobilfunknutzer-in-deutschland/.

Verhoef, P. C., Kannan, P. K., and Inman, J. J. From Multi-Channel Retailing to Omni-Channel Retailing. Introduction to the Special Issue on Multi-Channel Retailing. *Journal of Retailing*, 91(2):174–181, 2015.

Vlachos, P. A. and Vrechopoulos, A. Determinants of Behavioral Intentions in the Mobile Internet Services Market. *Journal of Services Marketing*, 22 (4):1–38, 2008.

Vlachos, P. A., Giaglis, G., Lee, I., and Vrechopoulos, A. P. Perceived Electronic Service Quality: Some Preliminary Results From a Cross-National Study in Mobile Internet Services. *International Journal of Human-Computer Interaction*, 27(3):217–244, 2011. doi: 10.1080/10447318.2011.537207. URL http://www.tandfonline.com/doi/abs/10.1080/10447318.2011.537207.

Vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., Cleven, A., Brocke, J. V., and Reimer, K. Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. In *Proceedings of the 17th European Conference on Information Systems (ECIS '17), Volume 9: 2206-2217*, Verona, Italy, 2009. ISBN 9788861293915.

Walmart. Walmart 2017 Annual Report, 2017a. URL http://s2.q4cdn.com/056532643/files/doc_financials/2017/Annual/WMT_2017_AR-(1).pdf.

Walmart. Our History, 2017b. URL http://corporate.walmart.com/our-story/our-history.

Walmart. Walmart: In-Store & Online Shopping. Easy Reorders, 2017c. URL https://itunes.apple.com/us/app/walmart-app-shopping-savings-catcher-more/id338137227?mt=8.

Webb, H. W. and Webb, L. A. SiteQual: An Integrated Measure of Web Site Quality. *Journal of Enterprise Information Management*, 17(6):430–440, 2004.

Webster, J. and Watson, R. T. Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, 26(2):xiii – xxiii, 2002.

Wiegers, K. and Beatty, J. *Software Requirements*. Microsoft Press, Redmond, 2. edition, 2013.

Wolfinbarger, M. and Gilly, M. .comQ: Dimensionalizing, Measuring, and Predicting Quality of the E-tail Experience. *Working Paper: Marketing Science Institute Report*, pages 1–42, 2002.

Wolfinbarger, M. and Gilly, M. C. eTailQ: Dimensionalizing, Measuring and Predicting Retail Quality. *Journal of Retailing*, 79(3):183–198, 2003.

Xing, Y., Grant, D. B., McKinnon, A. C., and Fernie, J. Physical Distribution Service Quality in Online Retailing. *International Journal of Physical Distribution & Logistics Management*, 40(5):415–432, 2010.

Yang, K. and Kim, H.-Y. Mobile shopping motivation: An Application of Multiple Discriminant Analysis. *International Journal of Retail & Distribution Management*, 40(10):778–789, 2012.

Yang, Z., Cai, S., Zhou, Z., and Zhou, N. Development and Validation of an Instrument to Measure User Perceived Service Quality of Information Presenting Web Portals. *Information and Management*, 42(4):575–589, 2005.

Yoo, B. and Donthu, N. Developing a Scale to Measure the Perceived Quality of An Internet Shopping Site (SITEQUAL). *Quarterly Journal of Electronic Commerce*, 2(1):1–19, 2001.

Zeithaml, V. A. Service Excellence in Electronic Channels. *Managing Service Quality: An International Journal*, 12(3):135–139, 2002.

Zeithaml, V. A., Parasuraman, A., and Malhotra, A. A Conceptual Framework for Understanding e-Service Quality: Implications for Future Research and Managerial Practice. *Working Paper: Marketing Science Institute*, pages 1–26, 2000.

Zhao, L., Lu, Y., Zhang, L., and Chau, P. Y. K. Assessing the Effects of Service Quality and Justice on Customer Satisfaction and the Continuance Intention of Mobile Value-added Services: An Empirical Test of a Multidimensional Model. *Decision Support Systems*, 52(3):645–656, 2012.

Zhu, H., Chen, E., Xiong, H., Cao, H., and Tian, J. Mobile App Classification with Enriched Contextual Information. *IEEE Transactions on Mobile Computing*, 13(7):1550–1563, 2014.

Zobel, J. *Mobile Business und M-Commerce: Die Märkte der Zukunft erobern*. München, Hanser, 2001.

Junior Management Science 4(3) (2019) 392-421

Junior Management Science

journal homepage: www.jums.academy

Bewertung der Übertragbarkeit von neuronalen Studienergebnissen auf einen Accounting-Kontext

Virginia Springer

Universität Stuttgart

Abstract

Erkenntnisse zu physiologischer und neuronaler Aktivität im Gehirn als grundlegende Auslöser ökonomischer Entscheidungen finden vermehrt Zuspruch in der betriebswirtschaftlichen Forschung. Ein noch junges Forschungsfeld stellt dabei das Neuroaccounting dar, welches zentrale verhaltens- und neurowissenschaftliche Erkenntnisse mit Grundsätzen des Accounting und der Psychologie verknüpft.

Das Ziel dieser Arbeit ist, einen umfassenden Überblick über das emergierende Feld des Neuroaccounting zugeben. Vor diesem Hintergrund wird dargestellt, inwiefern neurowissenschaftliche Erkenntnisse in einem Accounting-Kontext untersucht und übertragen werden und ob aufgrund kontextspezifischer Veränderungen Limitationen bei der Übertragung bestehen können. Hierzu werden die thematischen Zusammenhänge umfassend durch eine systematische Literaturanalyse aufbereitet.

Die daraus resultierenden Ergebnisse zeigen, dass in einem Accounting-Kontext weitgehend auf neurowissenschaftliche Erkenntnisse verwiesen werden, welche in einem Nicht-Accounting-Kontext gefunden wurden. Die kritische Betrachtung der Ergebnisse verdeutlicht, dass vor allem stichprobenrelevante Störvariablen sowie kontextspezifische psychologische Effekte zu Limitationen bei der Übertragung führen. Folglich ist eine Kontextabhängigkeit bei der bestehenden Übertragung neurowissenschaftlicher Erkenntnisse auf einen Accounting-Kontext gegeben, welche je nach Studiendesign im Kontext der eigentlichen Untersuchung zu berücksichtigen ist.

Keywords: neuroaccounting; decision making; contextuality; cognitive neuroscience; human brain.

1. Einleitung

1.1. Relevanz des Themas und Problemstellung

Entscheidungen sind ein wichtiges betriebswirtschaftliches Instrument in Unternehmen und werden von den jeweiligen Entscheidungsträgern fortlaufend in verschiedensten Unternehmensbereichen und -situationen getroffen.[1] Vor diesem Hintergrund sind sowohl strategische als auch operative Entscheidungen über relevante und umfassende Maßnahmen wichtige Elemente der nachhaltigen Koordination und Steuerung eines Unternehmens. Das sog. Accounting unterstützt Entscheidungsträger bei der Vorbereitung unternehmerischer Entscheidungen. Eine wichtige Funktion des Accounting besteht bspw. in der Gewinnung, Verarbeitung sowie der geeigneten Bereitstellung und Übermittlung von Informationen an Adressaten.[2] Informationen stellen die Grundlage für künftige Entscheidungen dar. Unternehmerisches Handeln im Accounting kann ferner als Kette von Entscheidungen betrachtet werden. Typische Entscheidungen im Accounting sind bspw. Entscheidungen über Investitionsausgaben oder Budgetvorgaben, welche als Grundlage für unternehmerischen Erfolg angesehen werden können. Aufgrund einer Kombination aus hoher Umwelt- und Aufgabenunsicherheit, sozialem Druck sowie Zeitdruck erhöht sich das Unsicherheitsniveau für die Entscheidungsträger erheblich. Diese Situation verleitet Entscheider dazu, sich auf vereinfachte Mechanismen und Heuristiken zu verlassen, um schnelle Entscheidungen treffen zu können.[3] Dies kann zu verschiedenen kognitiven Verzerrungen[4] führen, welche den Entscheidungsprozess und das daraus resultierende Verhal-

[1] Vgl. Jeschke (2017), S. 3.

[2] Vgl. Amshoff (1993), S. 272-274.

[3] Vgl. Haley und Stumpf (1989), S. 481; Schwenk (1984), S. 111f.

[4] Ein Beispiel für eine kognitive Verzerrung ist, wenn Entscheidungsträger die Aussagefähigkeit vorhandener Informationen sowie die eigenen Fähigkeiten nicht realistisch einschätzen (vgl. Barnes Jr (1984), S. 133.).

DOI: https://doi.org/10.5282/jums/v4i3pp392-421

ten stark beeinflussen.[5]

Durch die Beschaffung entscheidungsrelevanter Informationen soll dieser Unsicherheit entgegengewirkt werden. Die Entwicklung immer kürzerer Innovationszyklen und die effizientere Verbreitung von entscheidungsrelevanten Informationen soll eine schnellere und flexiblere Bereitstellung ermöglichen. Erreicht werden kann dies durch die neuartigen Informations- und Kommunikationstechnologien im digitalen Zeitalter.[6] Vorangegangene Forschungen nehmen sich dem Thema der daraus resultierenden Informationsüberlast an und liefern wichtige Erkenntnisse zur Beschaffung und Bereitstellung von Informationen. Hierbei werden jedoch zumeist nur die Eigenschaften von Informationen selbst und nicht die neuronalen Prozesse im menschlichen Gehirn während der eigentlichen Entscheidungsfindung untersucht. Schon Lashley verweist 1930 auf die Wichtigkeit der Neurowissenschaften bei der Erklärung des menschlichen Verhaltens. Er diskutiert dabei das Konzept der llocalization of function“[7] . Dieses betont, dass gewisse Neuronen im Gehirn für bestimmte kognitive Funktionen spezialisiert sind. Er stellt ausdrücklich fest, dass die eigentliche und endgültige Erklärung des Verhaltens oder der mentalen Prozesse in der physiologischen Aktivität des Körpers und insbesondere in den Eigenschaften des Nervensystems zu suchen ist.[8] Als eine mögliche Antwort auf den bestehenden Wunsch der Forschung, die „Black-Box“ des Gehirns besser verstehen zu können, hat sich seit Beginn des letzten Jahrzehnts die sog. Neuroökonomie an der Schnittstelle zwischen Wirtschaftswissenschaft, Psychologie sowie Neurowissenschaft etabliert. Diese liefert wertvolle Erkenntnisse zur Analyse menschlichen Verhaltens in ökonomischen Entscheidungssituationen.[9] Unterstützt wird die Neuroökonomie vor allem von der naturwissenschaftlichen Methodik der neuronalen Hirnforschung. Hier wird durch den Einsatz technischer Verfahren wie bspw. der funktionellen Magnetresonanztomographie aufzeigt, welche spezifischen Hirngebiete bei einer bestimmten Entscheidungssituation stimuliert und aktiviert werden.[10] Die dadurch resultierende Interdisziplinarität des Forschungsfeldes ermöglicht es Forschern, ein präziseres Verständnis über Vorgänge innerhalb des menschlichen Gehirns zu erlangen und kann so zur Erklärung von Kauf-, Investitions- oder auch Managemententscheidungen beitragen.

In betriebswirtschaftlichen Disziplinen, die sich mit der Konsumerforschung beschäftigen, werden mit zunehmender Intensität neuronale Erkenntnisse[11] und neurowissenschaftliche Methoden herangezogen.[12] Auch in der Accounting-Forschung wird diese Integration vermehrt gefordert.[13] Dabei bietet sich die Möglichkeit, aufschlussreiche Erkenntnisse sowohl für die wissenschaftliche Forschung als auch für die Praxis zu generieren.[14] In Accounting-Journalen werden vermehrt wissenschaftliche Arbeiten veröffentlicht, welche zur Stärkung der theoretischen Argumentation auf neurowissenschaftliche Studienergebnisse verweisen.[15] Waymire betont dabei, dass, um das Problem des kausalen Zusammenhangs bei Analysen im Accounting angehen zu können, direkte (neurowissenschaftliche) Erkenntnisse notwendig sind. Diese können aufzeigen, wie Prinzipien der Rechnungslegung im Gehirn verarbeitet werden.[16] Das sog. Neuroaccounting stellt dabei eine Bereichsdisziplin der Neuroökonomie und der verhaltensorientierten Accounting-Forschung dar, welche zur Erforschung von Problemen wie diesen beitragen kann. Beim Neuroaccounting werden zentrale verhaltenswissenschaftliche und neurowissenschaftliche Erkenntnisse mit Grundsätzen des Accounting und der Psychologie in Verbindung gebracht.[17] Daraus resultierende Ergebnisse können sich als nützlich erweisen, um bspw. Auswirkungen finanzieller Anreize auf die Informationsverarbeitung von Entscheidungsträger im Accounting besser zu verstehen.[18]

Durch den zunehmenden Trend der Verknüpfung verschiedenster wissenschaftlicher Disziplinen zur Entwicklung interdisziplinärer Forschungs- und Entwicklungsbereiche, kommt die Problemstellung auf, wie dieses Zusammenspiel wirtschaftswissenschaftlicher Fragestellungen mit neurowissenschaftlichen Methoden und Erkenntnissen in der Accounting-Forschung umgesetzt werden kann. Zudem soll geklärt werden, ob eine direkte Übertragung neurowissenschaftlicher Erkenntnisse, die in einem Nicht-Accounting-Kontext gefunden wurden, auf einen Accounting-Kontext möglich ist. Die Problemstellung der Arbeit orientiert sich dabei an den von Birnberg et al. identifizieren möglichen Forschungsentwicklungen des Neuroaccounting. Birnberg et al. beschäftigen sich in ihrem Essay „Is neuroaccounting waiting in the wings?“ von 2012 u.a. mit der Frage, welches Entwicklungspotential dem Neuroaccounting in der künftigen Forschung zugeschrieben werden kann.[19] Dabei identifizieren sie zwei mögliche Forschungsphasen. Laut Birnberg et al. ist davon auszugehen, dass das Neuroaccounting in seinem frühesten Stadium bestehende verhaltensökonomische Erkenntnisse im Accounting mit neurowissenschaftlichen Erkenntnissen in Verbindung bringen wird. Dies stellt Phase eins dar. Erst in einer zweiten (späteren) Phase des Neuroaccounting ist davon auszugehen, dass Methoden der Neurowissenschaften selbständig in der Bearbeitung von verhaltenswissenschaftlichen Frage- und Problemstellungen im Accounting eingesetzt werden.[20]

[5] Vgl. Tversky und Kahneman (1974), S. 1124.
[6] Vgl. Edmunds und Morris (2000), S. 17.
[7] Lashley (1930), S. 1.
[8] Vgl. Lashley (1930), S. 2-6.
[9] Vgl. Reimann und Weber (2011), S. 5.
[10] Vgl. Camerer (2008), S. 417.
[11] Neuronale Erkenntnisse sind im Rahmen dieser Arbeit die Folge neurowissenschaftlicher Forschungen. Nachfolgend werden somit die Begriffe neuronale Erkenntnisse und neurowissenschaftliche Erkenntnisse synonym verwendet.
[12] Vgl. Schilke und Reimann (2007), S. 247.
[13] Vgl. beispielsweise Oblak et al. (2018), Hecht et al. (2012).
[14] Vgl. Hofmann und Küpper (2011), S. 116.
[15] Ein Überblick hierzu ist in Kapitel 3.2 zu finden.
[16] Vgl. Waymire (2014), S. 2011.
[17] Vgl. Artienwicz (2016), S. 253.
[18] Vgl. Dickhaut et al. (2010), S. 223f.
[19] Vgl. Birnberg und Ganguly (2012), S. 7f.
[20] Vgl. Birnberg und Ganguly (2012), S. 8.

1.2. Zielsetzung und Forschungsfragen

Die geeignete Auswahl von Entscheidungsszenarien hat unter Berücksichtigung methodischer Bedingungen im Kontext der jeweiligen Forschungsarbeit zu erfolgen. Dies hat Auswirkungen auf die spätere Übertragbarkeit der Ergebnisse und ist elementar für eine fundierte wissenschaftliche Basis.[21] Aus dem thematischen Kontext ergibt sich daher die übergeordnete Zielsetzung, einen Überblick über das emergierende Feld des Neuroaccounting zu erhalten. Ausgehend von diesem übergeordneten Ziel sollen im Rahmen der vorliegenden Arbeit zwei Forschungsfragen beantwortet werden.

> Forschungsfrage 1: In wie weit wird in der Accounting-Forschung auf neurowissenschaftliche Erkenntnisse und Methoden zurückgegriffen?

Ausgehend von den in der ersten Teilfrage gefundenen Erkenntnissen, wird in der zweiten Forschungsfrage der Fokus stärker auf die qualitative Untersuchung der Ergebnisse hinsichtlich eines Accounting-Kontextes gelegt. Aus diesem Zusammenhang ergibt sich die folgende Fragestellung.

> Forschungsfrage 2: Sind neurowissenschaftliche Erkenntnisse aus einem Nicht-Accounting-Kontext auf einen Accounting-Kontext übertragbar oder bestehen aufgrund kontextspezifischer Veränderungen Limitationen, die sich auf die wissenschaftliche Übertragbarkeit auswirken?

Die dargestellten Forschungsfragen stecken den strukturellen Rahmen ab und bilden den Kern der vorliegenden Arbeit. Dabei soll die aufgeführte Zielsetzung sowohl quantitativ als auch qualitativ bearbeitet werden. Hierfür wird eine Literaturanalyse durchgeführt. Anschließend werden die Ergebnisse diskutiert. Um ein einheitliches Begriffsverständnis zu gewährleisten, wird nachfolgend dargestellt, wie im Rahmen dieser Arbeit der Begriff des Accounting-Kontexts zu verstehen ist.

Unter einem Kontext kann weitgehend der Zusammenhang der, abgesehen von der experimentell kontrollierten unabhängigen Variablen einwirkenden, relevanten Rahmenbedingungen verstanden werden. Dieser Zusammenhang kann das Verhalten eines Systems und den daraus resultierenden Output beeinflussen.[22] In einem Kontext selbst werden zumeist kontext-spezifische Entscheidungen getroffen. Nachfolgend wird darunter das Versuchsdesign einer experimentellen Studie verstanden, welches sich auf das Verhalten der Versuchsteilnehmer auswirkt. Im angloamerikanischen Raum wird für das Rechnungswesen in aller Regel der Begriff des Accounting verwendet. Dieser Begriff kann in Abhängigkeit des zu informierenden Empfängers grundsätzlich in Management- bzw. Cost-Accounting oder in Financial-Accounting aufgegliedert wird.[23] Der Untersuchungsbereich setzt sich vorrangig mit der internen Unternehmensrechnung auseinander und konzentriert sich auf die Leistungen des Rechnungswesens. Unter Leistung des Rechnungswesens sind in erster Linie die Entscheidungs- und Verhaltenssteuerungsfunktion[24] des Management Accounting zu verstehen. Diese dienen auf der einen Seite zur Deckung des sachlichen Koordinationsbedarfs von Entscheidungen. Auf der anderen Seite zielt die Verhaltenssteuerungsfunktion auf die personelle Koordination der Entscheidungsträger ab.[25] Infolgedessen ist nachfolgend unter Accounting stets das Management Accounting zu verstehen ist, sofern nicht anders gekennzeichnet.

Vor diesem Hintergrund ist unter einem Accounting-Kontext somit jenes (Entscheidungs-)Szenario in einer Studie zu verstehen, in welchem Entscheidungen anhand von immateriellen, relevanten, zuverlässigen und vergleichbaren Informationen getroffen werden. Hierbei besteht das langfristige Ziel der (Unternehmens-) Nutzenmaximierung. Die daraus resultierenden Ergebnisse dienen zusätzlich dem Zweck, die Leistung des Rechnungswesens in einer bestimmten Weise (egal ob in Form oder Inhalt) zu beeinflussen. Eine weitere Bedingung ist, dass die Beiträge in einem Accounting-Journal veröffentlicht wurden.[26] Darüber hinaus wird eine Stichprobe mit Finanz- oder Accounting-Fachbezug herangezogen oder Probanden, die die Rolle einer fiktiven Finanz- oder Accounting-relevanten Position einnehmen. Wird eine der genannten Kriterien nicht erfüllt, so handelt es sich nicht um einen Accounting-Kontext. Die Begriffe Szenario und Entscheidungsszenario werden nachfolgend synonym verwendet.

1.3. Aufbau der Arbeit

Die vorliegende Arbeit strukturiert sich in fünf Kapitel, die nachfolgend kurz erläutert werden. Im Anschluss an die Einleitung werden im zweiten Kapitel die theoretischen Grundlagen des Neuroaccounting als Teildisziplin verhaltensorientierter Wissenschaft dargestellt. Schwerpunktmäßig wird das verhaltensorientierte Accounting aufgezeigt, sowie die Teildisziplin des Neuroaccounting vorgestellt. Abschließend wird denjenigen neurowissenschaftlichen Messmethoden besondere Aufmerksamkeit gewidmet, welche bisher im Neuroaccounting zur Identifikation neuronaler Erkenntnisse eingesetzt wurden. Auf die Vorstellung der physikalischen Grundlagen der jeweiligen Methode folgt eine prägnante Diskussion der Stärken und Schwächen der einzelnen Verfahren. Im dritten Kapitel wird die dieser Arbeit zugrunde gelegte methodische Vorgehensweise, die sog. systematische Literaturanalyse, vorgestellt. Das Kapitel schließt mit einem Überblick über die gefundenen Ergebnisse ab. Hierbei liegt der

[21] Vgl. Hussy et al. (2012), S. 118; Renner et al. (2012), S. 80.
[22] Vgl. Bazire und Brézillon (2005), S. 38.
[23] Vgl. Drury (2012), S. 6.
[24] Vgl. Demski und Feltham (1976), S. 8-11.
[25] Vgl. Wall und Schröder (2009), S. 9f.
[26] Hierzu wird das VHB JOURQUAL 3 (2018) Teilranking Rechnungswesen als Bewertungsmaßstab herangezogen.

Fokus auf der Darstellung, wie in der aktuellen Accounting-Forschung auf neurowissenschaftliche Erkenntnisse zurückgegriffen wird. Zudem soll aufgezeigt werden, inwieweit neurowissenschaftliche Messmethoden im Accounting eingesetzt werden.

Im vierten Kapitel erfolgt eine Diskussion der in Kapitel drei gefunden Ergebnisse. Hierbei liegt der Fokus auf denjenigen Beiträgen, welche im Rahmen eines Accounting-Kontextes auf neurowissenschaftliche Erkenntnisse verweisen. Die identifizierten Beiträge werden vor allem vor dem Hintergrund der externen Validität von Untersuchungsergebnissen diskutiert. In Kapitel fünf werden die Ergebnisse der vorliegenden Arbeit zusammengefasst und ein Fazit gezogen. Abschließend werden Implikationen für zukünftige neurowissenschaftliche Forschungen im Accounting aufgezeigt.

2. Accounting unter neurowissenschaftlicher Perspektive

Das nachfolgende Kapitel soll einen Überblick über die Entwicklung der Accounting-Forschung geben. Der Fokus liegt hier auf der verhaltenswissenschaftlichen Teildisziplin des Neuroaccounting (Kapitel 2.1). Abschließend wird in diesem Kapitel eine Übersicht über diejenigen neurowissenschaftlichen Messmethoden gegeben, welche bisher in der Neuroaccounting-Forschung eingesetzt wurden (Kapitel 2.2).

2.1. Neuroaccounting als Teildisziplin verhaltensorientierter Wissenschaft

Obwohl die Bereitstellung und Berichterstattung von Finanzinformationen nach wie vor im Mittelpunkt der Rolle eines Accountants stehen, erhalten vor allem verhaltensbezogene Aspekte immer mehr Aufmerksamkeit.[27] Im verhaltensorientierten Accounting (VA), auch bekannt als „behavioral accounting research“, sind sowohl die Entscheidungsfunktion als auch die Verhaltenssteuerungsfunktion der internen Unternehmensrechnung Gegenstand extensiver Forschungsaktivitäten.[28] Das verhaltensorientierte Accounting untersucht in diesem Zusammenhang Zwecke, Aufgaben sowie Instrumente der betrieblichen Unternehmensrechnung aus einer entscheidungsorientierten Perspektive. Dabei steht „verhaltensorientiert“ für Verhaltensannahmen über die Adressaten von Informationen, welche von den standardwirtschaftstheoretischen Annahmen des Homo Oeconomicus abweichen.[29] Grund hierfür ist, dass diese Theorie von strenger Eigennützigkeit und unbeschränkter Rationalität, d.h. uneingeschränkten Kapazitäten bei der Informationsaufnahme sowie der fehlerfreien Verarbeitung dieser Informationen, ausgeht. Folglich vernachlässigen diese Annahmen die Existenz von kognitiven Verzerrungen. Das verhaltensorientierte Accounting ist, wie andere verhaltenswissenschaftliche Ansätze, interdisziplinär. Menschliches Verhalten wird nicht nur mit Hilfe ökonomischer, sondern primär auch unter Zuhilfenahme psychologischer sowie soziologischer Ansätze erklärt.[30] Aufgrund dieser interdisziplinären Betrachtungsweisen wird im VA nicht nur der Informationsinhalt der Unternehmensrechnung selbst, sondern auch die Darstellungsform und die Art und Weise der Übermittlung von Informationen untersucht.[31] Des Weiteren werden nicht nur die Ergebnisse von Entscheidungen, sondern auch insbesondere die jeweiligen Entscheidungsprozesse beleuchtet. Ziel des VA ist es dabei, die Auswirkungen der Ausgestaltungen der Unternehmensrechnung auf das menschliche Verhalten zu erklären und vorherzusagen.

Das sog. Neuroaccounting kann als junge Teildisziplin sowohl des verhaltensorientierten Accounting als auch der Neuroökonomie angesehen werden. Eine einheitlich etablierte Definition oder ein bestehendes Paradigma für das emergierende Forschungsfeld ist in der Literatur jedoch nicht zu finden. Zumeist wird lediglich darauf verwiesen, dass im Neuroaccounting zentrale verhaltenswissenschaftliche und neurowissenschaftliche Erkenntnisse mit Grundsätzen des Accountings und der Psychologie vereint werden. In Verbindung gebracht und untersucht wird dies durch bspw. die Messung der Hirnaktivität bei ökonomischen Entscheidungen.[32] Diese verbreitete multidisziplinäre Auffassung des Forschungsfeldes soll nachfolgend dargestellt werden. Hierbei wird erörtert, inwieweit jede dieser Disziplinen zum Feld des Neuroaccountings beiträgt. Andere Disziplinen, die hier nicht näher dargestellt werden, können ebenfalls in das Neuroaccounting einspielen. Im Rahmen dieser Arbeit liegt der Fokus jedoch auf den in Abbildung 1 dargestellten Teildisziplinen. Als Ausgangpunkt des Neuroaccountings ist das Accounting selbst anzuführen. Im angloamerikanischen Raum wird für das Rechnungswesen der Begriff des Accountings verwendet, welcher in Abhängigkeit des zu informierenden Adressatenkreis unterteilt werden kann. Im Allgemeinen kann das Accounting in Management- oder Cost-Accounting für das interne Rechnungswesen sowie in Financial-Accounting für das externe Rechnungswesen gegliedert werden.[33] Unter dem Begriff des Rechnungswesens sind dabei Konzepte und Verfahrensweisen zu verstehen, die eine ergebnisorientierte, quantitative Erfassung, Dokumentation, Aufbereitung sowie Auswertung ökonomischer Prozesse ermöglichen. Der Fokus liegt sowohl auf unternehmensinternen Prozessen als auch auf wirtschaftlich relevanten Prozessen des Unternehmens mit seiner Umwelt.[34] Vor allem in der angloamerikanischen Literatur ist die Abgrenzung zwischen interner und externer Rechnungslegung nicht eindeutig, da dem Management-Accounting der umfassendere Gesamtinhalt der Unternehmensrechnung, zugeordnet wird.[35] Vorrangig werden hier

[27] Vgl. Kim et al. (2012), S. 4.
[28] Vgl. Caplan (1966), S. 497-499.
[29] Vgl. Gillenkirch (2008), S. 128.
[30] Einen Überblick über psychologische Ansätze im VA geben Birnberg et al. (2007).
[31] Vgl. Jeschke (2017), S. 107.
[32] Vgl. beispielsweise Wang (2018), S. 329f; Lechman und Christiansen (2016), S. 253.
[33] Vgl. Drury (2012), S. 6.
[34] Vgl. Eisele und Knobloch (2014), S. 1.
[35] Vgl. Sweeny (1983), S. 467f.

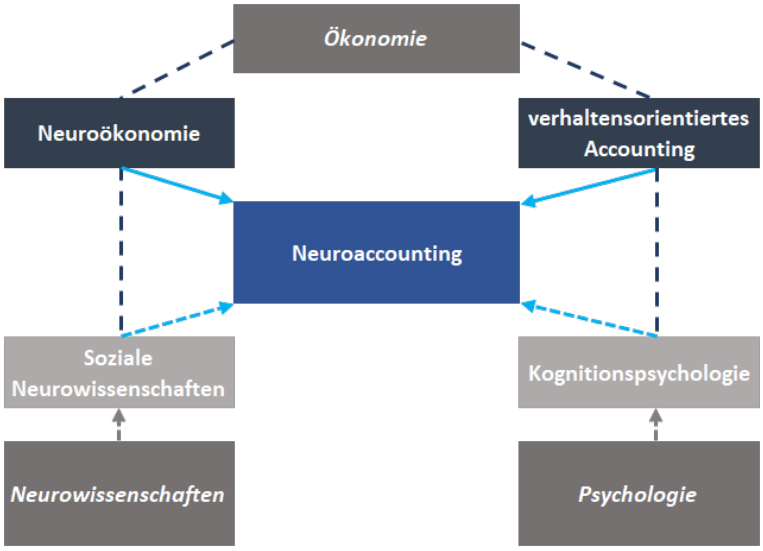


Abbildung 1: Theoretische Einordnung des Neuroaccountings; Eigene Darstellung.

Prinzipien und Grundsätzen der internen Unternehmensrechnung und die daraus resultierende Entscheidungs- und Verhaltenssteuerungsfunktion unter dem Begriff Accounting subsummiert. Beide Funktionen sind i.d.R. durch die Existenz von Unsicherheit gekennzeichnet. Die Entscheidungsfunktion ist dabei ein Informationsinstrument als Grundlage für Entscheidungen durch das Management. Diese Entscheidungen können aber auch im Accounting verankert sein. Anders als bei der Entscheidungsfunktion zielt die Verhaltenssteuerungsfunktion vorrangig auf die Beeinflussung von verschiedenen Entscheidungsträgern im Unternehmen ab.[36] Der Fokus liegt hier insbesondere auf Untersuchungen zur Informationsverarbeitung.

In diesem Zusammenhang bilden sowohl die normative als auch die deskriptive Entscheidungstheorie fundamentale Grundlagen der Unternehmensrechnung. Die normative Entscheidungstheorie geht von einem rationalen Handeln des Entscheidungsträgers aus und hat durch die Entwicklung von Entscheidungsregeln zum Ziel, das Treffen von rationalen Entscheidungen zu fördern. Rationale Entscheidungen sind dabei jene, welche unter den gegebenen Bedingungen zur größten Zielerreichung beitragen. Die deskriptive Entscheidungstheorie hingegen untersucht die Entstehung von Entscheidungen in der Realität. Hierbei wird vernachlässigt, ob eine Entscheidung tatsächlich rational oder aufgrund mangelnder Zielvorstellung emotional getroffen wird. Ziel ist somit, das realitätsnahe Entscheidungsverhalten zu untersuchen und empirisch zu belegen.[37] Als ein übergeordnetes Ziel der entscheidungstheoretischen Perspektive im Accounting kann folglich das systematische Treffen von Entscheidungen zur Lösung von Entscheidungsproblemen im Accounting und anderen Unternehmensbereichen angesehen werden. Reichmann sieht die Funktionen des Accountings zudem in der Analyse von Daten sowie deren Beurteilung und Kontrolle.[38] Für das Neuroaccounting bildet das Accounting selbst die wissenschaftliche Grundlage, die u.a. anhand des Einsatzes neurowissenschaftlicher Methoden den Entscheidungsprozess in der „Black Box“ des Gehirns transparenter darstellen soll. Dadurch sollen die Prozesse hinter der Entscheidungs- und Verhaltenssteuerungsfunktion im Accounting besser dargestellt werden. Eine weitere Teildisziplin des Neuroaccounting stellt die Neurowissenschaft dar. Diese beschäftigt sich mit der Erforschung des Gehirns, seiner Struktur und Funktion und damit, wie es das menschliche Verhalten beeinflusst.[39] Des Weiteren tragen die Neurowissenschaften dazu bei, transparent aufzuzeigen, wie verschiedene Gehirnstrukturen repariert und Gehirnfunktionen wiederhergestellt werden können, wenn diese beeinträchtigt sind. Die Neurowissenschaft selbst ist ein interdisziplinärer Ansatz, der sich aus Biologie, Chemie, Mathematik, Medizin, Physik und Psychologie zusammensetzt. Laut Mukamel et al. ist das oberste Ziel der neurowissenschaftlichen Forschung, den Wirkungsmechanismus des menschlichen Gehirns zu verstehen und dieses Verständnis zu nutzen, um Methoden zur Reparatur bei Fehlfunktionen zu entwickeln.[40] Die soziale Neurowissenschaft ist als Teildisziplin der Neurowissenschaft anzusehen. Unter sozialer Neurowissenschaft ist ein interdisziplinäres Forschungsfeld zu verstehen, welches untersucht, wie biologische Systeme soziale Prozesse und Verhaltensweisen implementieren können. Hierzu werden biologische Konzepte und Methoden zur Verfeinerung von Theorien sozialer Prozesse und Verhaltensweisen herangezogen.[41] Dabei werden vorrangig soziale Phänomene und Prozesse mit Hilfe neurowissenschaftlicher Forschungsinstrumente wie dem

[36] Vgl. Ewert und Wagenhofer (2008), S. 6-9.
[37] Vgl. Laux et al. (2014), S. 2.
[38] Vgl. Reichmann (1997), S. 3f.
[39] Vgl. Beugré (2018), S. 8.
[40] Vgl. Mukamel und Fried (2012), S. 511.
[41] Vgl. Cacioppo et al. (2007), S. 100.

Neuroimaging[42] untersucht.[43] Während der physiologische Aufbau des Gehirns in der Forschung schon lange bekannt ist, ist erst durch die bildgebenden Methoden der kognitiven Neurowissenschaften eine nicht-invasive[44] Messung der Aktivitäten einzelner Hirnregionen möglich.[45] Diese Erkenntnisse unterstützen das Neuroaccounting dahingehend, dass eigene soziale Verhalten und das anderer besser zu verstehen, sowie Studiensituationen zu kontrollieren und eine direkte Lokalisation stimulierter Signale zu ermöglichen.

Eine weitere wichtige Disziplin des Neuroaccounting ist die Neuroökonomie. Die Neuroökonomie bezieht sich auf die Untersuchung der Auswirkungen des Gehirns auf ökonomische Entscheidungen.[46] Die Neuroökonomie ist ein Teilgebiet der Verhaltensökonomie und der experimentellen Ökonomie und befasst sich mit den neuronalen Grundlagen von menschlichen Verhaltensweisen. Aufgrund der interdisziplinären Aufstellung dieses Forschungsfeldes werden neurowissenschaftliche Messtechniken mit ökonomischen Methoden und Theorien kombiniert. Vor diesem Hintergrund wird untersucht, wie das menschliche Gehirn Entscheidungen im betriebswirtschaftlichen und sozialen Kontext generiert und verarbeitet.[47] Eines der primären Ziele der Neuroökonomie ist es, die nicht beobachtbaren, subjektiven Aspekte der Entscheidungsfindung in wirtschaftlichen Situationen zu erforschen.[48]

Camerer et al., welche als Pioniere der Neuroökonomie angesehen werden können, behaupten, dass im engeren Sinne bei jeder Untersuchung einer wirtschaftlichen Tätigkeit das menschliche Gehirn einbezogen werden soll.[49] Laut Braeutigam kann an dieser Stelle betont werden, dass der menschliche Verstand der Treiber allen wirtschaftlichen Handelns ist und vor allem verhaltenswissenschaftliche Studien elementare kognitive Prozesse im Gehirn nicht ignorieren können.[50] Die Neuroökonomie stellt insbesondere die übliche ökonomische Annahme in Frage, dass die Entscheidungsfindung ein einheitlicher Prozess ist. Man geht vielmehr davon aus, dass die Entscheidungsfindung sowohl durch das Zusammenspiel von automatischen als auch kontrollierten Prozessen angetrieben wird.[51] Volk und Köhler argumentieren zudem, dass die Neuroökonomie Forschern helfen kann, bessere Theorien über die zugrundeliegenden Ursachen für beobachtbares Verhalten in einem bestimmten Kontext zu entwickeln, sowie präzisere Tests ihrer Theorien zu erstellen.[52] Diese Argumentation lässt sich ebenfalls auf das Neuroaccounting übertragen, denn die Neuroökonomie untersucht Phänomene, die für das Neuroaccounting ebenfalls von Relevanz sind. So untersuchen Neuroökonomen bspw. Themen wie Entscheidungsfindung, Unsicherheit, Vertrauen und Kooperation sowie emotionale Grundlagen des menschlichen Verhaltens. Hierbei werden wichtige Erkenntnisse gefunden, die ebenso die Entscheidungs- und Verhaltenssteuerungsfunktion im Accounting verbessern können. Die Neuroökonomie kann folglich als übergreifende Disziplin angesehen werden, die die Grundsätze der Ökonomie aus einer neurowissenschaftlichen Perspektive untersucht. Das Neuroaccounting kann als Teilbereich der Neuroökonomie angesehen werden. Des Weiteren können wichtige bestehende neuroökonomische Erkenntnisse zu bspw. dem Zusammenspiel von affektiven und kognitiven Reaktionen zur transparenteren Darstellung von Entscheidungsprozessen im Accounting genutzt werden.

Als letzte wichtige Disziplin im Kontext des Neuroaccountings ist die Psychologie anzuführen. Hierbei liegt der Fokus vor allem auf der Kognitionspsychologie, auch kognitive Psychologie genannt. Sie untersucht in erster Linie, wie Menschen Informationen mental darstellen und verarbeiten. Als solches werden vor allem psychische Vorgänge wie Wahrnehmung, Lernen, Gedächtnis, Denken, Problemlösen und die Entscheidungsfindung selbst ergründet.[53] Laut Poldrack besteht das Ziel der kognitiven Psychologie darin, die zugrunde liegende mentale Architektur des Menschen zu verstehen, welche die kognitiven Funktionen unterstützt.[54] Der starke Fokus auf Kognitionen (d.h. die Informationsverarbeitung selbst) trägt zur Entwicklung von Theorien über die Rolle der Kognition bei der Entscheidungsfindung unter Unsicherheit bei. Darüber hinaus hat diese auch zur Entwicklung wichtiger Theorien wie der Prospekttheorie[55] und der Theorie kognitiver Verzerrungen bei der Entscheidungsfindung[56] beigetragen. Folglich unterstützt die kognitive Psychologie das Neuroaccounting dahingehend, die mentale Struktur von Gedanken und menschlichem Verhalten durch psychologische Erkenntnisse besser darzustellen.

Nachfolgend ist unter Neuroaccounting somit jenes multidisziplinäre Forschungsfeld zu verstehen, das zur Erforschung Accounting-spezifischer Problemstellung verschiedenste Erkenntnisse aus Psychologie, Neurowissenschaft und Neuroökonomie nutzt. Somit kann gewährleistet werden, dass Problem- und Fragestellung des VA aus verschiedenen Perspektiven beleuchtet werden. Durch den Verweis auf neurowiss. Erkenntnisse und den Einsatz neurowiss. Methoden soll ferner eine Verbesserung der Hauptfunktionen des Management Accountings ermöglicht werden. Zielsetzung des Neuroaccounting ist hier, durch die Integration der genannten Forschungsdisziplinen, ökonomisch relevantes Verhal-

[42] Neuroimaging bezeichnet im Allgemeinen die Erstellung von Bildern des Nervensystems im menschlichen Gehirn durch bildgebende Verfahren (vgl. Brammer (2009), S. 389.).

[43] Vgl. Lieberman (2007), S. 260.

[44] Nicht-invasiv ist ein Begriff aus der Medizin und bezeichnet Verfahren, bei denen Messgeräte oder damit verbundene Katheter nicht in den Körper des Untersuchten eindringen (vgl. Kligman (2006), S. 4.).

[45] Vgl. Roth (2003), S. 124-127.

[46] Vgl. beispielsweise Glimcher (2004); Camerer et al. (2005).

[47] Vgl. beispielsweise Zak (2004); Braeutigam (2005); Fehr et al. (2005).

[48] Vgl. Camerer et al. (2007), S. 138.

[49] Vgl. Camerer et al. (2004), S. 555.

[50] Vgl. Braeutigam (2005), S. 355.

[51] Vgl. Loewenstein et al. (2008), S. 647.

[52] Vgl. Volk und Köhler (2012), S. 523.

[53] Vgl. Sternberg (1981), S. 1181.

[54] Vgl. Poldrack (2006), S. 59.

[55] Hierzu siehe Kahneman und Tversky (1979).

[56] Hierzu siehe Tversky und Kahneman (1974).

ten besser zu verstehen und aus den daraus resultierenden Erkenntnissen Implikationen für die Accounting-Forschung abzuleiten. Letztendlich soll somit der ökonomische Entscheidungsprozess im Accounting transparenter dargestellt werden.

2.2. Überblick über relevante neurowissenschaftliche Messmethoden

In Anlehnung an die von Riedl et al.[57] vorgeschlagene Untergliederung von neurowissenschaftlichen Messmethoden werden die nachfolgend dargestellten Verfahren anhand des jeweiligen zu messenden Nervensystems gegliedert. In diesem Zusammenhang werden zwei Kategorien unterschieden: Psychophysiologische Verfahren (z.B. Eye-Tracking), welche Aktivitäten und Stimulationen des peripheren Nervensystems untersuchen, sowie bildgebende Verfahren (z. B. Elektroenzephalografie, funktionelle Magnetresonanztomographie), welche Aktivitäten und Stimulationen des zentralen Nervensystems[58] untersuchen (siehe Tabelle 2).
[59]

Im Rahmen dieser Arbeit werden lediglich diejenigen neurowissenschaftlichen Messmethoden erörtert und dargestellt, welche auch Anwendung in der Accounting-Forschung finden (siehe hierzu Kapitel 3). Der Schwerpunkt dieses Kapitels liegt auf den sog. bildgebenden Verfahren, da vor allem diese in den vergangenen Jahren stark an Bedeutung für die Neurowissenschaften gewonnen haben.[60]

2.2.1. Psychophysiologische Verfahren

Eine der bekanntesten psychophysiologischen Methoden zur Messung von Stimulationen im zentralen Nervensystem ist dabei das sog. Eye-Tracking.[61] Bei diesem nicht-invasiven Verfahren werden in erster Linie Blickbewegungen eines Probanden aufgezeichnet und als physiologische Reaktion erfasst.[62] Blickbewegungen setzen sich aus Fixationen und Sakkaden zusammen. Fixationen sind wenige Millisekunden andauernde Phasen, während denen der Blickverlauf verharrt und das Auge beginnt, (An-)Reize aus der Umwelt aufzunehmen. Sakkaden hingegen sind Augenbewegungen, die zwischen verschiedenen Fixationen stattfinden, um den Blick neu auszurichten.[63] Bei einigen Eye Tracking-Studien wird zusätzlich der Durchmesser der Pupille gemessen, da vorangegangene Studien gezeigt haben, dass geweitete Pupillen ein Anzeichen für erhöhte Aufmerksamkeit sind.[64] Bei sogenannten kopfbasierten Systemen tragen Probanden zudem einen am Kopf befestigten Sensor. Aufgrund neuster Entwicklungen in der Computertechnik ist es möglich, durch in einem Bildschirm hinterlegte Infrarotsensoren die Blickbewegung eines Probanden aufzuzeichnen.[65] Dies lässt folglich die Untersuchung einer großen Stichprobengruppe zu.

Psychophysiologische Verfahren wie das Eye-Tracking sind im Vergleich zu bildgebenden Verfahren einfach in ihrer Anwendung, wodurch Ergebnisse schnell und relativ kostengünstig verfügbar sind.[66] Zudem sind die für eine Eye Tracking-Untersuchung notwendigen Geräte relativ leicht transportierbar. In Folge dessen kann eine bessere Validität der Ergebnisse auch außerhalb eines Labors erzielt werden.[67] Zudem findet Eye-Tracking vermehrt auch in betriebswirtschaftlichen Teildisziplinen Anwendung.[68] Schwankungen der Raumtemperatur oder verschiedene Tageszeitformen sind mögliche Störvariablen, die während des Eye-Trackings auftreten und die Ergebnisse beeinflussen können.[69] Zudem ist durch den Einsatz von psychophysiologischen Verfahren wie bspw. dem Eye-Tracking keine direkte Messung neuronaler Aktivität im Gehirn des Probanden möglich. Anstelle dessen wird ausgehend von körperlichen Reaktionen des Probanden, d.h. dem Bewegungsverlauf der Augen, auf spezifische Gehirnfunktionen bzw. daraus resultierende physiologische Prozesse geschlossen.[70]

2.2.2. Bildgebende Verfahren

Bildgebende Verfahren können in sogenannte elektrophysiologische und metabolische Verfahren untergliedert werden.[71] In diesem Zusammenhang messen elektrophysiologische Verfahren die direkte Hirnaktivität, wohingegen bei metabolischen Verfahren eine indirekte Messung von Substanzen, welche mit neuronalen Aktivitäten verbunden sind, erfolgt.[72] Ergebnisse der Berechnungen eines bildgebenden Verfahrens sind zumeist eine für die neurale Aktivität nahezu vollständige und bildliche Darstellung. In diesem Kontext wird zwischen räumlicher und zeitlicher Auflösung unterschieden. Die Fähigkeit, Details klar aufzunehmen, bezeichnet man als räumliche Auflösung.[73] Aus je mehr Pixel sich dabei eine Aufnahme zusammensetzt, desto höher ist die räumliche Auflösung und umso genauer ist eine Lokalisierung neuronaler Aktivität möglich. Die zeitliche Auflösung gibt zudem an, wie groß der zeitliche Abstand zwischen zwei Aufnahmen ist.[74] Dabei gilt: Je höher die zeitliche Auflösung ist, desto kürzer ist der Abstand zwischen den Aufnahmen und desto direkter ist die Messung der Aktivität bzw. des Aktionspotentials von Neuronen.

[57] Vgl. Riedl und Léger (2016), S. 47.
[58] Das Rückenmark und Gehirn bilden das zentrale Nervensystem. Zum peripheren Nervensystem hingegen zählen alle neuronalen Strukturen, die sich außerhalb des zentralen Nervensystems befinden (vgl. David und Aguayo (1981), S. 931.).
[59] Eigene Darstellung.
[60] Toga et al. (2012) betonen, dass „Neuroimaging has been and still is one of the most important tools available to study the brain in vivo“ (S. 323).
[61] Vgl. Duchowski (2017), S. xi.
[62] Vgl. Shimojo et al. (2003), S. 1321.
[63] Vgl. Rayner (1998), S. 372f.
[64] Vgl. Pieters und Wedel (2007), S. 225.
[65] Vgl. Kenning (2014), S. 142.
[66] Vgl. Camerer et al. (2005), S. 14.
[67] Vgl. Kenning (2014), S. 142.
[68] Vgl. Stewart et al. (2016), S. 117.
[69] Vgl. Lee et al. (2007), S. 200.
[70] Vgl. Duchowski (2017), S. 3.
[71] Vgl. Raichle (1998), S. 765.
[72] Vgl. Kenning (2014), S. 86.
[73] Vgl. Kimberley und Lewis (2007), S. 671.
[74] Vgl. Kimberley und Lewis (2007), S. 671.

Tabelle 1: Überblick über verwendete neurowissenschaftliche Methoden im Accounting

	Funktion und Methode	Stärken	Schwächen
Psychophysiologische Verfahren (peripheres Nervensystem)			
Eye-Tracking	- Bewegung der Pupille als Anzeichen für neuronale Aktivität - Befestigung von Sensoren an Kopf des Probanden.	- Weit verbreitet, - nicht-invasiv, - einfache Anwendung, - relativ günstig, - Validität.	- Keine direkte Messung, da lediglich Rückschluss auf neuronale Aktivität, - anfällig für Störvariablen der Umgebung.
Bildgebende Verfahren (zentrales Nervensystem)			
Elektrophysiologische Verfahren			
Elektroenzephalografie	- Erkennung von elektrischen Potentialunterschieden an der Kopfhautoberfläche, die durch neuronale Aktivität verursacht werden - Proband trägt Reihe von Elektroden am Kopf.	- lange Anwendungshistorie (wiss. anerkannt). - nicht-invasiv, - ausgezeichnete zeitliche Auflösung, - Stichprobengröße.	- Aktivität einzelner Neuronen zu schwach, da erst Signal von Neuronengruppen an Kopfoberfläche messbar (begrenzte räumliche Auflösung).
Metabolische Verfahren			
funktionelle Magnetresonanztomographie	- Erkennung von Magnetfeldverzerrungen, die sich aus Veränderungen der kortikalen Aktivität oder Gewebeschäden ergeben können - Untersuchung neuronaler Aktivität in Form von Veränderungen des Sauerstoffgehalts (BOLD-Kontrast) - Proband in MR-Scanner.	- Weit verbreitet, - nicht-invasiv, - Quelle des Signals kann sehr genau lokalisiert werden (ausgezeichnete räumliche Auflösung).	- Teuer, - Validität, - neueste Technologie nicht für alle verfügbar, - schlechte zeitliche Auflösung bei Signaländerungen, - „reverse inference“.

Zu den gebräuchlichsten elektrophysiologischen und ältesten neurowissenschaftlichen Verfahren zählt die sog. Elektroenzephalografie (EEG).[75] Die EEG ist ein nicht-invasives Verfahren, bei welchem zur Messung von elektrischen Spannungsschwankungen an der Schädeloberfläche Elektroden am Kopf des Probanden angebracht werden.[76] Aufgrund der hohen zeitlichen Auflösung lassen sich dynamisch-kognitive Prozesse in wenigen Millisekunden abbilden.[77] Zudem ist die EEG relativ kostengünstig und es können im Gegensatz zu anderen bildgebenden Verfahren auch größere Stichprobengrößen untersucht werden.[78] Zusätzlich sind aufgrund der relativ einfachen Datenanalyse die Ergebnisse schnell verfügbar.[79]

Darüber hinaus kann im Rahmen von EEG-Studien eine höhere Realitätsnähe der experimentellen Situation erreicht werden. Denn ein EEG ist geräuscharm und lässt Bewegungen des Probanden zu. Da ein EEG relativ einfach räumlich transportiert werden kann, ist die Durchführung von Experimenten nicht zwingend nur im Labor möglich.[80] Die räumliche Auflösung ist jedoch niedriger als bei anderen bildgebenden Verfahren, da lediglich das Signal großer Neuronen-Gruppen an der Kopfoberfläche messbar ist. In diesem Fall ist aber zu beachten, dass durch die Anzahl der Elektroden die räumliche Auflösung erheblich beeinflusst werden kann.[81] Des Weiteren erfolgt durch den Einsatz eines EEGs lediglich die Messung von Aktivitäten in den äußersten Gehirnschichten, wodurch eine Lokalisierung von Aktivitäten in tieferen Schichten des Gehirns (z.B. Nervenzellen) nicht möglich ist.[82] Dennoch können durch den Einsatz der EEG wertvolle Erkenntnisse über bspw. das Präferenzverhalten von Menschen gewonnen werden.[83]

Ein wichtiges Verfahren zur Messung der metabolischen Aktivität ist die sog. funktionelle Magnetresonanztomographie (fMRT), welche in der neuronalen Forschung zu der am weitesten verbreiteten Messmethode gehört.[84] Durch den Einsatz einer nicht-invasiven fMRT ist eine direkte Beobachtung bestimmter Hirnregionen, die während eines Experiments aktiviert und stimuliert werden, möglich. Hierbei werden Magnetfeldverzerrungen, die sich aus der Veränderung der kortikalen Aktivität[85] oder Gewebeschäden ergeben, gemessen.[86] Vor allem die starke Kombination aus räumlich-zeitlicher Auflösung und anatomischer Abdeckung rechtfer-

[75] Vgl. Luck (2014), S. 4.
[76] Vgl. Michel und Murray (2012), S. 372.
[77] Vgl. Kable (2011), S. 70.
[78] Vgl. Ariely und Berns (2010), S. 288.
[79] Vgl. Lorensen und Dickson (2003), S. 55.
[80] Vgl. Salmelin und Baillet (2009), S. 1754.
[81] Vgl. Ariely und Berns (2010), S. 288.
[82] Vgl. Camerer et al. (2005), S. 12.
[83] Telpaz et al. setzten bspw. in ihrer Studie ein EEG zur Prognose der zukünftigen Produktwahl verschiedener Konsumenten ein (Telpaz et al. (2015, S. 512).
[84] Kable (2011) zeigt dabei auf, dass ca. 70 Prozent aller empirisch-neurowissenschaftlichen Studien zur Entscheidungsforschung auf ein fMRT zurückgreifen (S. 76).
[85] Kortikale Aktivität bezeichnet neuronale Aktivität in der sog. Hirnrinde, welche auch Kortex genannt wird (vgl. Chen et al. (2008), S. 2374.).
[86] Vgl. Schröger (2011), S. 61.

tigt den häufigen Einsatz der fMRT in den Neurowissenschaften, so Kable.[87] Die vorherrschende Variable, die bei Einsatz eines fMRT untersucht wird, ist der sog. „blood oxygenation level dependent“, auch BOLD-Kontrast genannt. Hierbei wird der Sauerstoffgehalt in den roten Blutkörperchen untersucht, welcher sich aus einer Veränderung der Durchblutung im Gehirn als Reaktion auf verschiedene Reize ergibt.[88] Dies beeinflusst maßgeblich die Qualität des Bildsignals. Vor diesem Hintergrund ist zu berücksichtigen, dass aufgrund der zeitlichen Differenz zwischen dem Auftreten dynamischer Veränderungen im Blutfluss und der eigentlichen elektrochemische Reaktion in den Neuronen, die zeitliche Auflösung reduziert werden kann.[89] Zudem führen Dulleck et al. an, dass aufgrund der Größe der Technologie selbst (MRT-Scanner) die simultane Beobachtung einer großen Anzahl von Teilnehmern schwierig ist.[90] Dies kann zu Einschränkungen bei der Größe der Stichprobe und Versuchsbedingungen selbst führen. Ferner werden Möglichkeiten der Feldforschung eingeschränkt, da sich die Versuchsperson bei Durchführung der Studie im Scanner befinden muss. Dies kann die Validität und somit die Generalisierbarkeit der Ergebnisse beeinträchtigen. Des Weiteren sind die Geräte mit hohen Kosten verbunden und die Verfügbarkeit in betriebswirtschaftlichen Teildisziplinen begrenzt.[91]

Ein Ziel bildgebender Verfahren besteht darin, nachzuweisen, dass sich zwei Stimuli in einem gegebenen Prozess unterscheiden, indem man Unterschiede im abgeleiteten Signal untersucht. Die logische Konsequenz, die sog. „forward inference“, schließt daher vom kognitiven Prozess auf das eigentliche Signal. Der Umkehrschluss vom Signal auf den kognitiven Prozess, die sog. „reverse inference“, ist ebenfalls weit verbreitet in der Forschungspraxis und wird kritisch diskutiert. Laut Poldrack ist die kausale Schlussfolgerung in Form einer „reverse inference“ aufgrund der geringen (lokalisatorischen) Signifikanz von kognitiven Prozessen logisch unzulässig , sofern diese nicht statistisch überprüft wurde.[92] Grund hierfür ist, dass letztlich bestimmten Strukturen im Gehirn so gut wie jede kognitive Bezeichnung und Rolle zugeordnet werden kann. In der Forschungsliteratur wird daher empfohlen, fMRT- und EEG-Verfahren stets in Verbindung mit anderen (neurowiss. Mess-) Methoden zu verwenden, um eine ausreichende Validität der Ergebnisse zu gewährleisten.[93]

3. Neurowissenschaftliche Methoden und Erkenntnisse im Accounting

Im folgenden Kapitel wird die methodische Vorgehensweise der vorliegenden Arbeit erläutert. Hierzu wird zunächst die grundsätzliche Forschungsmethodik dargestellt (Kapitel 3.1). Die Darstellung beginnt mit der Entwicklung der Zielsetzung, beschreibt die einzelnen Schritte der Literaturbeschaffung und schließt mit der Konzeption des Hauptteils. Abschließend erfolgt eine kurze Zusammenfassung der Ergebnisse der Literaturanalyse (Kapitel 3.2). Nachfolgend werden die Begriffe Beitrag, Paper und Veröffentlichung synonym verwendet.

3.1. Vorstellung und Anwendung der methodischen Vorgehensweise

Zur Schaffung einer aussagekräftigen und relevanten Datengrundlage wurde eine systematische Literaturanalyse durchgeführt. Diese ist Teil einer qualitativen Forschungsvorgehensweise.[94] Hierbei werden die zu einer konkreten Fragestellung relevanten sowie zugänglichen Veröffentlichungen nach im Voraus genau definierten (Such-)Kriterien ermittelt, ausgewählt und anschließend diskutiert[95] Die Intention dabei ist, einen möglichst umfassenden Überblick über die vorherrschende Literatur zu erlangen, welche die Basis zur qualitativen Diskussion der dieser Arbeit zugrunde liegenden Forschungsfragen darstellt.[96] Die wissenschaftliche Vorgehensweise der vorliegenden Literaturanalyse erfolgte in Anlehnung an vom Brocke et al.[97] und untergliedert sich primär in vier Phasen. Die Fünf-Schritt-Vorgehensweise von vom Brocke et al. wurde in den Phasen eins und vier berücksichtigt (grau unterlegt) sowie um weitere notwendige Zwischenschritte in Phase zwei und drei ergänzt (siehe Abbildung 2).

Die Phasen eins und zwei dienen der grundlegenden Definition relevanter Suchkriterien (Stichwörter) sowie der eigentlichen Durchführung der Suche. In der dritten Phase erfolgt die Auswahl und eindeutige Darstellung aller für die Diskussion der Forschungsfragen geeigneten Veröffentlichungen. Phase vier rundet die Vorgehensweise anschließend mit einer Diskussion der gefunden Ergebnisse ab. Diese ist in Kapitel 4 zu finden. Zusätzlich empfehlen vom Brocke et al. im letzten Schritt die Erstellung einer Forschungsagenda, welche hier als Implikationen für zukünftige Forschungen zu verstehen ist. Diese sind in Kapitel 5 zu finden. Obwohl die Darstellung der verwendeten Vorgehensweise noch keine unmittelbaren Antworten auf die zu bearbeitenden Forschungsfragen liefert, ist die Strukturierung der Vorgehensweise dennoch ein wichtiger erster Schritt. Die Schaffung eines einheitlichen Verständnisses ist unerlässlich, um hierauf aufbauend die Beantwortung der Forschungsfragen zu ermöglichen.[98]

Eine Eingrenzung des Umfangs sowie eine Konzeptualisierung der Thematik wurde in Anlehnung an die Taxonomie von Cooper vorgenommen (siehe Abbildung 3). Dabei erfolgt eine Kategorisierung bezüglich des Umfangs und der Zielsetzung der Literaturanalyse nach [1] Fokus, [2] Ziel, [3]

[87] Vgl. Kable (2011), S. 67.
[88] Vgl. Logothetis und Pfeuffer (2004), S. 1524.
[89] Vgl. Balthazard und Thatcher (2015), S. 94f.
[90] Vgl. Dulleck et al. (2011), S. 118.
[91] Vgl. Kenning (2014), S. 88; Reimann und Weber (2011), S. 611f.
[92] Vgl. Poldrack (2006), S. 59f.
[93] Vgl. Butler et al. (2016), S. 544.
[94] Vgl. Töpfer (2012), S. 241.
[95] Vgl. Webster und Watson (2002), S. xvi.
[96] Vgl. Gash (2000), S. 3.
[97] Vgl. Vom Brocke et al. (2009), S. 9.
[98] Vgl. Vom Brocke et al. (2009), S. 9.

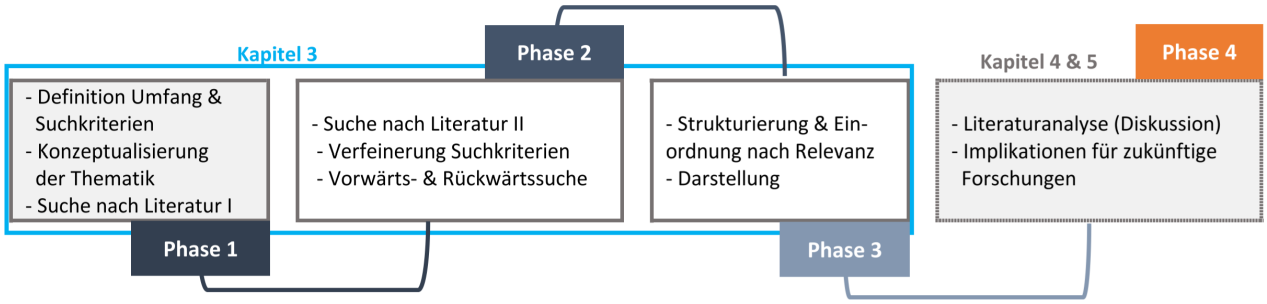


Abbildung 2: Methodisches Vorgehen der Arbeit; Eigene Erstellung.

Organisation, [4] Perspektive, [5] Zielgruppe und [6] Abdeckung.[99] In den Kategorien der Charakteristika Fokus, Ziel, Organisation und Zielgruppe sind Mehrfachzuordnungen zulässig. Die Subkategorien Perspektive und Abdeckung hingegen schließen sich wechselseitig aus. Hier ist vorwegzunehmen, dass im weiteren Verlauf keine geeigneten Monographien oder sonstige Bücher identifiziert wurden. Daher beziehen sich die nachfolgenden Erläuterungen ausschließlich auf Journalbeiträge.

Die Kategorie Fokus bezeichnet denjenigen Teil einer Veröffentlichung, welcher ausgehend von einer breiten Recherche analysiert und diskutiert werden soll. Hier liegt der Schwerpunkt auf den Forschungsmethoden sowie den zur (theoretischen) Argumentation eingesetzten Theorien. Ziel dieser Literaturanalyse ist eine Integration, welche auf eine Aufarbeitung und neutrale und historische Darstellung der aktuellen Forschungssituation abzielt. Darüber hinaus sollen Implikationen für zukünftige Forschungen aufgezeigt werden. Als Zielgruppe sind in erster Linie Wissenschaftler anzusehen, wobei grundsätzliche die gefundenen Ergebnisse auch für die die anderen Zielgruppen von Relevanz sein können. Zudem wird eine repräsentative Abdeckung dieser systematischen Literaturanalyse angestrebt, da nach einer vollumfänglichen Suche eine Beschränkung auf das thematisch Relevante sowie auf wissenschaftlich hochwertige Literatur erfolgt. Als Grundlage wurden hierfür Beiträge in Journalen, welche ein Peer-Review-Verfahren[100] verwenden, hinsichtlich verschiedener Stichwörter untersucht. Unter Berücksichtigung des vom Verband Deutscher Hochschullehrer für Betriebswirtschaft e.V. geführten Journalrankings JOURQUL3 wurden daher insgesamt sieben Journale der Rating-Kategorien A+ sowie A im Teilranking Rechnungswesen identifiziert (siehe Tabelle **??**). Zur Gewährleistung einer vollumfänglichen Identifikation geeigneter Beiträge wurden die frei verfügbare Websuchmaschine Google Scholar und die Fachdatenbank EBSCO herangezogen. Zusätzlich wurde eine direkte Suche auf den Internetseiten der Journale vorgenommen.

In einer ersten und zweiten Phase wurden zunächst alle geeigneten Beiträge gesammelt. Unter Eignung wird hier das eindeutige Auffinden anhand von im Vorhinein definierten Suchkriterien verstanden. Diese Kriterien stellen eine Auswahl an Stichwörtern dar, welche das Erkennen relevanter Beiträge ermöglichen sollen. Im Rahmen dieser Arbeit wurde eine iterative Stichwortsuche vorgenommen.[101] Ein Überblick über alle verwendeten Stichwörterkombinationen ist in Tabelle **??** zu finden. Hierbei ist anzumerken, dass die Stichwortsuche ebenfalls dazu beiträgt, die Parameter der Suche sowie den Umfang der Forschungsarbeit selbst zu konkretisieren.[102] Dabei wurde in den Suchoptionen der jeweiligen Journale und Datenbanken in allen Teilen eines Journals (Einleitung, Theorie, Methode, Ergebnisse, Diskussion, Zusammenfassung) nach den aufgeführten Stichworten gesucht. Dies wird in Tabelle **??** als „alle Felder"-Suche bezeichnet. Da es sich bei den vorliegenden Journalen um englischsprachige Werke handelt, erfolgte die Suche der für die Thematik dieser Arbeit relevanten Beiträge ebenfalls durch englischsprachige Suchbegriffe.

Die vorgenommene Konkretisierung soll dabei helfen, einen ersten Überblick über das Ausmaß der bestehenden Literatur zu erhalten, welcher als Vorbereitung für die eigentliche Klassifizierung und Auswahl in Phase drei dient. Mittels einer Verfeinerung der in der ersten Phase gewählten Stichwörter und einer gezielten vorwärts- und rückwärtsgerichteten Recherche in Phase zwei soll eine repräsentative Abdeckung der Literatur gewährleistet werden. Bei einer vorwärts-gerichteten Recherche werden diejenigen Beiträge ermittelt, welche die betrachtete Veröffentlichung zitiert haben. Bei der rückwärts-gerichteten Recherche hingegen wurden diejenigen Beiträge untersucht, die von der betrachteten Veröffentlichung zitiert wurden.[103] In der dritten Phase dieser systematischen Literaturanalyse erfolgte eine Strukturierung und Einordnung nach Relevanz der in der ersten und zweiten Phase identifizierten Beiträge. Als relevant werden dabei jene Beiträge eingestuft, welche auf neurowissen-

[99] Vgl. Cooper (1988), S. 104.

[100] Peer-Review Verfahren bezeichnen i.d.R. eine kritische Begutachtung wissenschaftlicher Texte durch unabhängige (dritte) Experten, welche prüfen, ob u.a. dargelegte wissenschaftliche Texte wissenschaftlichen Grundprinzipien entsprechen (vgl. Starck (2018), S. 1.).

[101] Die verwendeten Stichwörter wurden ausgehend von dem Fachbegriff des Neuroaccounting festgelegt und sequentiell um weitere Stichworte der relevanten Teildisziplinen ergänzt und angepasst. In der zweiten Phase der Suche liegt der Fokus der Stichwörter verstärkt auf Begriffen der Neurowissenschaften.

[102] Vgl. Baker (2000), S. 222.

[103] Vgl. Webster und Watson (2002), S. xvi.

Charakteristika		Kategorien			
1	Fokus	Forschungsergebnisse	Forschungsmethoden	Theorien	Anwendungen
2	Ziel	Integration	Kritik		Zentrale Themen
3	Organisation	Historisch	Konzeptionell		Methodisch
4	Perspektive	Neutrale Darstellung		Kritik	
5	Zielgruppe	Spezialisierte Wissenschaftler	Wissenschaftler	Praxis	Allgemeinheit
6	Abdeckung	Vollständig	Vollständig und selektiv	Repräsentativ	Zentral

Abbildung 3: Taxonomie nach Cooper; Entnommen aus Cooper (1988), S. 109.

Tabelle 2: Übersicht über betrachtete Journale, Datenbanken und verwendete Stichwörter; Eigene Darstellung.

Journal	Ranking	Suche	Treffer	Verwendete Stichwörter
Accounting Review	A+	"alle Felder"	13	• Phase 1:
Journal of Accounting and Economics	A+	"alle Felder"	1	neuroeconomics accounting
Journal of Accounting Research	A+	"alle Felder"	6	accounting neuroscience
Accounting, Organizations and Society	A	"alle Felder"	21	neuroscience
Contemporary Accounting Research	A	"alle Felder"	6	neuraoaccounting
Review of Accounting Studies	A	"alle Felder"	3	neuroscience accounting
Management Accounting Research	A	"alle Felder"	7	neuro accounting
			Σ: 57 Beiträge	neuro accounting finance
Datenbank		**Suche**		accounting science neuro
EBSCO		"alle Felder"		• Phase 2:
Google Scholar		"alle Felder"		neural science accounting
				neural accounting
				neural
				neuron
				acconting decision neuroeconomic
				accounting neuro economics

schaftliche Erkenntnisse (NWE) verweisen oder selbst neurowissenschaftliche Methoden (NWM) anwenden. Zudem wurde untersucht, ob ein Accounting-Kontext vorliegt.

3.2. Ergebnisse

Als Folge der Kategorisierung und Einordnung nach den in Kapitel 3.1 definierten Kriterien konnten insgesamt 57 Beiträge durch die angewandte Stichwortsuche ermittelt werden. Ein Überblick der ermittelten Beiträge ist in Tabelle 3 zu finden.

Von den 57 identifizierten Beiträgen wurden 22 Veröffentlichungen als thematisch nicht für diese Arbeit relevant eingestuft. Grund hierfür ist, dass nicht auf neurowissenschaftliche Studienergebnisse oder neurowissenschaftliche Messmethoden zurückgegriffen wurde. Des Weiteren werden in einigen Veröffentlichungen der Einsatz sowie Potenziale und Probleme künstlicher Intelligenz in der Accounting-Forschung thematisiert. Diese beschäftigt sich mit maschinell erstellten neuronalen Netzen und hat daher bei der Stichwortsuche ebenfalls Treffer ergeben. Folglich wurden somit 34 Beiträge zur näheren Kategorisierung identifiziert. Darüber hinaus verweisen 20 Beiträge lediglich auf neurowissenschaftliche Grundlagen oder führen auf, dass sich Disziplinen wie die Neuroökonomie oder das Neuroaccounting in der Forschung etablieren. Final werden somit insgesamt 14 Beiträge aus fünf verschiedenen Journalen identifiziert und weiterverarbeitet. Dabei ist zu erkennen, dass allein fünf dieser Beiträge in dem Journal „The Accounting Review" veröffentlicht wurden. In Anlehnung an die Gliederung nach Birnberg et al., verweisen neun Beiträge auf NWE. Ferner verwenden zudem sechs Beiträge selbst NWM.

3.2.1. Einsatz neurowissenschaftlicher Methoden im Accounting

Die final identifizierte Anzahl von 14 Beiträgen zeigt, dass die Anzahl der Veröffentlichungen, die auf NWE zurückgreift oder selbst NWM verwendet, gering ist. Führt man nun die von Birnberg et al. möglichen zwei Phasen des Neuroaccounting an, wird anhand der Ergebnisse ebenfalls deutlich, dass häufiger auf NWE verwiesen wird (neun Beiträge) als das selbst NWM eingesetzt werden (sechs Beiträge). Vor allem die Anzahl der Beiträge, welche neurowissenschaftliche Messmethoden in der Accounting-Forschung einsetzen, ist

Tabelle 3: Zusammenfassung und Klassifizierung der identifizierten Beiträge; Eigene Darstellung.

				Auswahl		
Journal	Treffer	[Phase 1]	[Phase 2]	Verweis NWE	Einsatz NWM	Irrelevant
Accounting Review	13	6	7	2	3	8
Journal of Accounting and Economics	1	-	1	-	-	1
Journal of Accounting Research	6	2	4	2	1	3
Accounting, Organizations and Society	21	9	12	1	1	19
Contemporary Accounting Research	6	5	1	2	-	4
Review of Accounting Studies	3	2	1	-	-	3
Management Accounting Research	7	4	3	2	1	4
Gesamt	57	28	29	9	6	42

bisher überschaubar. Dabei können drei der identifizierten Beiträge unter der Kategorie Entscheidungen subsummiert werden. Drei weitere Beiträge können der Kategorie Verhaltenssteuerung zugeordnet werden. Die Einteilung der Kategorien erfolgt in Anlehnung an die Funktionen des Management Accounting und orientiert sich an den untersuchten Thematiken. Die zeitliche Verteilung der Veröffentlichungen reicht dabei von 1997 bis 2018.[104] Die Studie selbst wird bei allen Beiträgen in einem Accounting-Kontext durchgeführt. Ein Überblick hierzu ist in Tabelle 4 zu finden. Diese Beiträge werden nachfolgend kurz dargestellt.

Die Studien von Barton et al. und Farrell et al. sind dabei die ersten und bislang einzigen Veröffentlichungen im Accounting, die grundlegende Fragen des Accountings durch den Einsatz einer fMRT untersuchen. Barton et al. untersuchen in diesem Zusammenhang, wie eine Gewinnausschüttung zu messbaren physiologischen Aktivitäten im Gehirn führt. Farrell et al. untersuchen in ihrer Studie, wie die Intensität affektiver Einflüsse auf Investitionsentscheidungen variieren, wenn ein leistungsorientierter Vertrag oder ein Festlohnvertrag verwendet wird. Eskenazi et al. nutzen zudem bisher als einzige die EEG, um Auswirkungen von sozialem Druck am Arbeitsplatz auf die Berichterstattung zu untersuchen. In Summe ist der Einsatz bildgebender Verfahren in der Accounting-Forschung mit drei Beiträgen bis dato überschaubar.

Der Einsatz psychophysiologischer Verfahren, hier speziell das Eye-Tracking, ist ebenfalls mit nur drei Veröffentlichungen überblickbar. Hunton et al., Chen et al. und Fehrenbacher et al. stellen bislang die einzigen Forschungsarbeiten im Accounting dar, die zur Untersuchung der Informationsverarbeitung diese Methodik einsetzen. Hunton et al. untersuchen dabei den Zusammenhang zwischen der Genauigkeit in der Berichterstattung und der dabei angewandten kognitiven Suchstrategie der Augen. Chen et al. untersuchen zudem Gestaltungsmöglichkeiten einer Balanced-Scorecard zur (positiven) Beeinflussung der Entscheidungsfindung. Fehrenbacher et al. untersuchen hingegen den Zusammenhang zwischen dem Spill-Over Effekt[105] und der subjektiven Leistungsbeurteilung.

3.2.2. Verweis auf neurowissenschaftliche Erkenntnisse im Accounting

Nach Abschluss der Literaturanalyse wurden sechs Beiträge identifiziert, welche zur (theoretischen) Hypothesenentwicklung oder Argumentation neurowissenschaftliche Erkenntnisse aufführen. Diese lassen sich ebenfalls den Kategorien Entscheidungen oder Verhaltenssteuerung zuordnen. Die zeitliche Verteilung der Veröffentlichungen reicht dabei von 2001 bis 2018.[106] Die jeweiligen Beiträge werden nachfolgend kurz vorgestellt. Zusätzlich wird aufgezeigt, wo und wie auf neurowissenschaftliche Studienerkenntnisse zurückgegriffen wird. Ein Überblick hierzu ist in Tabelle 5 zu finden. An dieser Stelle ist vorwegzunehmen, dass bei den nachfolgend aufgeführten neurowissenschaftlichen Studien kein Accounting-Kontext vorliegt (kAK). Lediglich ein Beitrag verweist auf neurowissenschaftliche Erkenntnisse, bei welchen Entscheidungsszenarien wie das Ultimatum-Spiel und das Gefangenendilemma eingesetzt werden. Ausgehend davon wird diesen neurowissenschaftlichen Studien ein betriebswirtschaftlicher Kontext (bwK) zugeordnet. Zusätzlich ist anzumerken, dass die Beiträge von Oblak et al. und Hecht et al. darauf hinweisen, dass die Robustheit der gefundenen Ergebnisse durch tiefere Untersuchungen mittels neurowissenschaftlicher Messmethoden für künftige Forschungen zu empfehlen ist.[107]

Des Weiteren sind drei Beiträge von Peecher et al., Waymire und Zingales anzuführen, welche ökonomisch-politische Probleme im Kontext des Financial Accounting im Rahmen einer theoretischen Argumentation untersuchen. Dabei verweisen alle drei Beträge auf neurowissenschaftliche Erkenntnisse, welche jedoch nicht näher diskutiert werden. Zudem erfolgt kein direkter Bezug zu der Thematik

[104] In diesem Zusammenhang ist zu berücksichtigen, dass ein zeitlicher Versatz zwischen Einreichung und Veröffentlichung eines Beitrages bestehen kann.

[105] Unter Spillover-Effekten versteht man im Allgemeinen Übertragungseffekte, bei welchen zunächst als unwichtig erscheinende Ereignisse/Zustände Auswirkungen auf andere Ereignisse/Zustände haben (vgl. Nilsson et al. (2017), S. 573.).

[106] Auch hier ist der zeitliche Versatz zwischen Einreichung und Veröffentlichung eines Beitrages zu berücksichtigen.

[107] Vgl. Oblak et al. (2018), S. 45; Hecht et al. (2012), S.567.

Tabelle 4: Übersicht über den Einsatz neurowissenschaftlicher Messmethoden im Accounting; Eigene Darstellung.

Autor(en)	Zitation	Thema	Methodik	Teilnehmer	Ergebnis
			Entscheidungen		
Fehrenbacher et al. (2018)	1	Spill-Over-Effekt in der subjektiven Leistungsbewertung	Eye-Tracking	123 Accounting-Studenten	Subjektive Leistungsbewertung von Vorgesetzten wird von objektiven Leistungsniveaus unterer Hierarchieebenen beeinflusst Spill-Over-Effekt wird reduziert, wenn Einzelpersonen subjektive Informationen mit einem bewussteren Entscheidungsmodus verarbeiten als mit einem intuitiveren Entscheidungsmodus.
Chen et al. (2016)	18	Ausgestaltung einer Balanced-Scorecard zur Entscheidungsbeeinflussung	Eye-Tracking	60 Teilnehmer [Ø Berufserfahrung: 5 J.; Durchschnittsalter: 25 J.]	Manager, die mehr Wert auf strategisch vernetzte Leistungskennzahlen legen, treffen eher richtige Entscheidungen.
Farrell et al. (2014)	41	Einfluss Gestaltung des Vergütungsdesign auf Investitionsentscheidung	fMRT	Experiment 1: 30 Studenten [Rechtshänder; männl.] Experiment 2: 96 Studenten [weibl. und männl.]	Im Vergleich zu Festverträgen verringern leistungsbasierte Verträge den Anteil wirtschaftlich kostspieliger Entscheidungen.
Eskenazi et al. (2016)	17	Zusammenhang neurobiologische Eigenschaften von Controllingexperten und treuhänderische Verpflichtungen [Unternehmensethik]	EEG	29 professionelle Controller [5 weibl. / 24 männl. Durchschnittsalter: 34,9 J.]	Sozialer Druck durch BU Managern beeinflusst späteres Berichtsverhalten.
Barton et al. (2014)	26	Zusammenhang Preisbildung von Erträgen und Verarbeitung von Belohnungen im Gehirn	fMRT	35 nicht-professionelle Investoren [8 weibl. / 27 männl. Durchschnittsalter: 28 J.]	Asymmetrische Reaktion auf positive und negative Ertragsüberraschungen
Hunton und McEwen (1997)	232	Zusammenhang Genauigkeit m Berichtswesen und kognitiven Suchstrategie	Eye-Tracking	60 professionelle Finanzanalysten [20 weibl. / 40 männl. Ø Berufserfahrung: 8,9 J.; Durchschnittsalter: 35,7 J.]	Genauere Analysten nutzen eine direkte Suchstrategie nach Informationen während weniger genaue Analysten eine sequentielle Suchstrategie nutzen. Motivationsanreize verstärken die Tendenz von Analysten zu optimistischen Gewinnprognosen

des jeweiligen Beitrages. Aufgrund des Verweises auf NWE werden diese drei Beiträge aus Vollständigkeitsgründen an dieser Stelle angeführt, in der weiteren Diskussion in Kapitel 4 jedoch nicht berücksichtigt.

Tabelle 6 zeigt einen Überblick darüber, wie auf neurowissenschaftliche Erkenntnisse im Accounting verwiesen wird. Hierbei werden nur diejenigen neurowissenschaftlichen Studien des jeweiligen Accounting-Beitrages aufgeführt, von welchen ausgehend Kausalschlüsse im Laufe der theoretischen Argumentation gezogen werden. Des Weiteren werden nur die neurowissenschaftlichen Beiträge aufgeführt, welche im Rahmen einer experimentellen Studie gewonnen wurden und nicht auf einer theoretischen Argumentation beruhen.[108]

Die aus den Beiträgen von Oblak et al. und Kida et al.

[108] Da Blay/Douthit/Fulmer (Tabelle 5) lediglich auf neurowissenschaftliche Erkenntnisse verweisen, die im Rahmen einer theoretischen Argumentation aufgezeigt wurden, wird der Beitrag in Tabelle 6 nicht aufgeführt und

Tabelle 5: Übersicht über Veröffentlichungen im Accounting, die auf neurowissenschaftliche Erkenntnisse verweisen; Eigene Darstellung.

Autor(en)	Thema	Neurobezug	Methodik	Teilnehmer	Ergebnis
			Entscheidungen		
Oblak et al. (2018)	Einfluss von kognitiven Rahmen durch Anreizdesigns sowie Einfluss von Fairness der Vergütung auf Entscheidungen über Risiko und Aufwand	Einleitung	Experiment	100 Studenten [74 weibl. / 26 männl. Ø Berufserfahrung: 2,1 J.; Ø Alter: 23,5 J.]	Fairness der Vergütungszahlung im Vergleich zu anderen wirkt stärker auf Entscheidungen als Anreizdesign.
Kida et al. (2001)	Auswirkungen affektiver Reaktionen auf Investitionsentscheidungen	Theoretischer Hintergrund; Hypothesenentwicklung	Experiment	114 Manager [Ø Berufserfahrung: 8 J.]	Manager berücksichtigen Finanzdaten und affektive Reaktionen zur Bewertung des Nutzens einer Anlagealternative.
			Verhaltenssteuerung		
Blay et al. (2019)	Zusammenhang Ehrlichkeit und Reporting	Theoretischer Hintergrund	Experiment	Studenten[4 Runden mit je 10 bis 28 Teilnehmer]	Präferenz für Ehrlichkeit entsteht aus dem Wunsch, negative Auswirkungen von Verstößen gegen soziale Normen zu vermeiden.
Jia et al. (2014)	Zusammenhang Gesichtsstruktur von CEOs und Fehlberichterstattung	Einleitung; Theoretischer Hintergrund	Verhältnismessung im Gesicht	Auswahl von Bildern von 1.136 CEOs von S&P1500-UN im Jahr 2009	Positive Korrelation zwischen männlichen Gesichtsattributen und Auftreten von Fehlberichten.
Hecht et al. (2012)	Zusammenhang Spill-Over-Effekt und Anreize	Theoretischer Hintergrund; Hypothesenentwicklung	Experiment	58 Studenten	Negative Auswirkungen von Teilanreizen auf die unbelohnte Aufgabenerfüllung kann teilweise durch einen Spill-Over-Effekt gemildert werden.
Maas et al. (2011)	Einfluss von sozialen Präferenzen auf die Nutzung von Ermessensspielraum durch Manager bei der Bewertung. [Unternehmensethik]	Hypothesenentwicklung; Verweis Fußnote	Experiment	126 Studenten [42 weibl. / 84 männl. Ø Alter: 20,4 J.]	Bereitschaft, kostspielige Informationen anzunehmen, steigt, je stärker die Gesamtleistung eines Teams zur Bewertung der individuellen Leistung herangezogen wird.
			Sonstige		
Peecher et al. (2013)	Überblick über bestehenden Audit-Regularien zur Verbesserung der Auditing-Qualität	Argumentation im Text	theoretische Argumentation	-	Regulatorische Verantwortlichkeiten von Auditoren erfolgen in Form von Strafen und nicht von Belohnungen. Diese hängen von den Auditergebnissen und nicht von Attributen des Urteils der Auditoren ab.
Waymire (2009)	Rolle des Accounting in Unternehmen	Implikationen für zukünftige Forschungen	theoretische Argumentation	-	Forderung nach Accounting besteht schon im frühsten Stadium eines Unternehmens.
Zingales (2009)	Reform zu Corporate Governance und Schutz nicht-professioneller Anleger	Argumentation im Text	theoretische Argumentation	-	Vertrauensverlust der Anleger kann durch Reform in Finanzmärkten verbessert werden.

in der weiteren Diskussion in Kapitel 4 ausgeschlossen. Eine Gesamtübersicht inklusive der Verweise, die nicht aus experimentellen Studien gezogen werden, ist im Anhang zu finden.

Tabelle 6: Überblick über Verweise auf neurowissenschaftliche Erkenntnisse im Accounting; Eigene Darstellung. GA = Gruppen-Ableitung

Autor(en)	#	Neuro-Bezug	Untersuchte Variable	Funktion im Text	Szenario	Probanden
			Entscheidungen			
Oblak et al. (2018)	1	Hughes et al. (2014)	Motivation	Bezug, GA	kAK	Menschen
		Treadway et al. (2015)	Motivation	Bezug, GA	kAK	Menschen [gesund vs. krank]
		Walton et al. (2006)	Motivation	Bezug, GA	kAK	Ratten; Affen
		Wardle et al. (2012)	Koffeingehalt und Aufwand	Bezug, GA	kAK	Menschen
		Cocker et al. (2012)	Koffeingehalt und Aufwand	Bezug, GA	kAK	Ratten
		Hosking et al. (2014a)	geistige Anstrengung und Aufwand	Bezug, GA	kAK	Ratten
		Hosking et al. (2014b)	geistige Anstrengung und Aufwand	Bezug, GA	kAK	Ratten
		Salamone et al. (1994)	Dopamin und Entscheidungsverhalten	Bezug, GA	kAK	Ratten
Kida et al. (2001)	119	Bechara et al. (1996)	Wahrnehmung Strafen und Belohnungen	Bezug, Ableitung	kAK	Menschen [gesund vs. krank]
			Verhaltenssteuerung			
Jia et al. (2014)	84	Carré et al. (2010)	Testosteron und Aggression	Bezug, GA	kAK	Menschen [m]
		Geniole et al. (2012)	Testosteron und Aggression	Bezug, GA	kAK	Menschen [m]
		Penton-Voak et al. (2006)	Attribute und Persönlichkeitsurteile	Bezug, GA	kAK	Menschen
		Verplaetse et al. (2007)	Erkennungsmodus und Betrug	Bezug, GA	kAK	Menschen
		Apicella et al. (2014)	Testosteron und Risikoaversion	Bezug, GA	kAK	Menschen [m]
		Mehta et al. (2008)	Testosteron und sozialer Status	Bezug, GA	kAK	Menschen [m]
		Pound et al. (2008)	Testosteron und Wettbewerbsverhalten	Bezug, GA	kAK	Menschen [m]
		Zuckerman und Kuhlman (2000)	Persönlichkeit und Risikoaversion	Bezug, GA	kAK	Menschen
		Eisenegger et al. (2010)	Testosteron und Fairness	Bezug, GA	kAK	Menschen [w]
		Wright et al. (2012)	Testosteron und Entscheidungen	Bezug, GA	kAK	Menschen [w]
Hecht et al. (2012)	27	Pochon et al. (2002)	Motivation und Kognition	Bezug, Ableitung	kAK	Menschen
		Taylor et al. (2004)	Motivation	Bezug, Ableitung	kAK	Menschen
		Beck et al. (2010)	Motivation	Bezug, GA	kAK	Menschen
		Engelmann et al. (2009)	Motivation und Aufmerksamkeit	Bezug, GA	kAK	Menschen
		Locke und Braver (2008)	Motivation	Bezug, GA	kAK	Menschen
Maas et al. (2011)	63	Sanfey et al. (2003)	Emotion und Kognition	Bezug Fußnote, Ableitung	bwK	Menschen
		Singer et al. (2006)	Empathische Reaktion	Bezug Fußnote, Ableitung	bwK	Menschen
		De Quervain et al. (2004)	Empathische Reaktion [insb. Schmerz]	Bezug Fußnote, Ableitung	bwK	Menschen

gewonnenen Erkenntnisse lassen sich der Kategorie Entscheidung zuordnen. Oblak et al. untersuchen dabei den Einfluss von kognitiven Rahmen durch Anreizdesigns sowie den Einfluss von Fairness in Bezug auf die Vergütung bei Entscheidungen über Risiko und Aufwand.[109] In der Einleitung und theoretischen Hinführung zur Thematik verweisen sie auf verschiedene neurowissenschaftliche Erkenntnisse, die sich grundsätzlich alle mit den Themen Motivation und Aufwand beschäftigen. Kida et al. hingegen untersuchen die Auswirkungen affektiver Reaktionen auf Investitionsentscheidungen.[110] Hierbei stützen sie sich sowohl beim theoretischen Hintergrund als auch der Hypothesenentwicklung auf neurowissenschaftliche Studienerkenntnisse von Bechara et al. zum psychologischen Phänomen eines Affekts.[111] Hierbei ist anzumerken, dass Oblak et al. darauf hinweisen, dass die zukünftige Forschung, ausgehend von den gefundenen Erkenntnissen, eine umfassende kognitive Theorie über den Risikoaufwand entwickeln kann. Dies kann jedoch nur unter der Einbeziehung von neurowissenschaftlichen Forschungsmethoden erfolgen.[112]

Die aus den Beiträgen von Jia et al., Hecht et al. und Maas et al. gewonnenen Erkenntnisse können ferner der Kategorie Verhaltenssteuerung zugeordnet werden. Jia et al. untersuchen hierbei den Zusammenhang zwischen der Gesichtsstruktur von CEOs und der Fehlberichterstattung.[113] Die eingesetzte Methode ist dabei die sog. „facial width-to-height“ Messung. Bei dieser Methode wird der Abstand zwischen den Wangenknochen sowie die Höhe zwischen Oberlippe und dem höchsten Punkt des Augenlieds auf verschiedenen Fotografien bestimmt.[114] Jia et al. stützen sich in der Einführung und theoretischen Entwicklung auf neurowissenschaftliche Erkenntnisse, welche explizit den Zusammenhang zwischen dem Hormon Testosteron und verschiedenen Persönlichkeitsmerkmalen und -attributen im Rahmen neurowissenschaftlicher Studien untersuchen.

In der Kategorie Verhaltenssteuerung können ferner Hecht et al. angeführt werden. Diese untersuchen den Zusammenhang zwischen dem Spill-Over Effekt und Anreizen.[115] Hecht et al. verweisen primär auf neurowissenschaftliche Studienerkenntnisse, welche exekutive Funktionen[116] vor dem Hintergrund verschiedener Motivationszustände untersuchen. Zudem führen Hecht et al. an, dass sie bei der Untersuchung des Spillover-Effekts aufgrund von technischen Einschränkungen nicht direkt die physiologische Reaktion auf Teilanreize untersuchen. Infolgedessen weisen sie in einer Fußnote aus, dass durch bspw. den Einsatz einer fMRT in zukünftigen Forschungen eine direkte Messung physiologischer Aktivitäten zur weiteren Validierung der Ergebnisse vorgenommen werden kann.[117]

Abschließend kann der Beitrag von Maas et al. in der Kategorie Verhaltenssteuerung angeführt werden. Maas et al. untersuchen in ihrer Veröffentlichung von 2012 den Einfluss von sozialen Präferenzen auf die Nutzung von Ermessensspielraum durch Manager bei der Leistungsbewertung.[118] In einer Fußnote verweisen sie auf neuroökonomische Erkenntnisse von Sanfey et al., Singer et al. und De Quervain et al., welche untersuchen, wie wahrgenommene Ungerechtigkeit und daraus resultierende empathische Reaktionen in unterschiedlicher neuraler Aktivität resultieren.

Grundsätzlich ist zu erkennen, dass die Quantität der neurowissenschaftlichen Verweise je nach Beitrag stark variiert. Während Kida et al. zur theoretischen Argumentation lediglich einen Beitrag von Bechara et al. aufführen, weisen Oblak et al. und Jia et al. eine deutlich höhere Verweisquote auf. Hierbei ist jedoch zu berücksichtigen, dass Unterschiede in der Verweisvorgehensweise bestehen. Wird ausgehend von einer neurowissenschaftlichen Studie eine theoretische Ableitung vorgenommen, so handelt es sich um einen Bezug mit anschließender Ableitung. Wird jedoch ausgehend von mehr als einer neurowissenschaftlichen Studie eine theoretische Ableitung vorgenommen, so ist dies als Gruppen-Ableitung zu verstehen. Zudem kann zwischen kontextspezifischen Verweisen oder Kausalschlüssen unterschieden werden. In diesem Zusammenhang ist anzumerken, dass lediglich Jia et al. und Hecht et al. untersuchungsspezifische Verweise aufführen. Untersuchungsspezifische Verweise bezeichnen hierbei die kontextspezifischen Ergebnisse aus der neurowissenschaftlichen Studie, die sich aus den untersuchten Variablen ergibt. Die verbleibenden aufgeführten Beiträge verweisen lediglich oberflächlich durch Kausalschlüsse auf die angeführten Studien und nehmen keine untersuchungsspezifischen Verweise vor.

Darüber hinaus ist zu erkennen, dass bis auf Maas et al. alle anderen aufgeführten Accounting-Beiträge auf neurowissenschaftliche Erkenntnisse im Fließtext verweisen und zur Stärkung der theoretischen Argumentation heranziehen. Zudem verweisen alle in Tabelle 6 aufgeführten Beiträge auf neurowissenschaftliche Erkenntnisse, die in keinem Accounting-Kontext gewonnen wurden. Lediglich Maas et al. verweisen auf neurowissenschaftliche Erkenntnisse, die in einem betriebswirtschaftlichen Kontext untersucht und gewonnen wurden. Den in Maas et al. verwiesenen Studien von Sanfey et al., Singer et al. und De Quervain et al. wird ein betriebswirtschaftliches Entscheidungsszenario zugeschrieben, da hier Entscheidungsszenarien wie das

[109] Vgl. Oblak et al. (2018), S. 35.
[110] Vgl. Kida et al. (2001), S. 477.
[111] Der Begriff Affekt bezieht sich auf eine Reihe von emotionalen Zuständen. Dabei wird angenommen, dass affektive Reaktionen eine Vielzahl von emotionalen Reaktionen - wie Wut, Frustration und Freude - die mit einer positiven oder negativen Valenz assoziiert sind, abbilden (vgl. Fiske und Taylor (2013). S. 342.).
[112] Vgl. Oblak et al. (2018), S. 45.
[113] Vgl. Jia et al. (2014), S. 1195.
[114] Vgl. Jia et al. (2014), S. 1205.
[115] Vgl. Hecht et al. (2012), S. 564.
[116] Unter exekutiven Funktionen sind hierbei jene geistigen Funktionen zu verstehen, mit denen Individuen das eigene Verhalten und Denken unter Berücksichtigung bestehender Bedingungen ihrer Umgebung steuern. Hierzu zählen bspw. Zielsetzungen sowie Entscheidungen über Prioritäten (vgl. Burgess und Simons (2004), S. 212.).
[117] Vgl. Hecht et al. (2012), S. 567.
[118] Vgl. Maas et al. (2011), S. 617.

Ultimatum-Spiel[119] sowie das Gefangenendilemma[120] eingesetzt werden. Diese sind in der experimentellen Ökonomieforschung etablierte Entscheidungsszenarien, die häufig zur Untersuchung von ökonomischem Entscheidungsverhalten in Bezug auf Zusammenarbeit und Fairness eingesetzt werden.[121] Des Weiteren wird in den Beiträgen von Sanfey et al. und De Quervain et al. explizit ausgewiesen, dass diese dem Forschungsstrang der Neuroökonomie zugeordnet werden.[122] Ferner kann angemerkt werden, dass in keiner der verwiesenen neurowissenschaftlichen Studien Probanden mit Finanz- oder Accounting-spezifischen Erfahrungen untersucht werden. Auffällig hierbei ist, dass von Oblak et al. auf neurowissenschaftliche Studienerkenntnisse verwiesen wird, welche mit nicht-humanen Primaten erzielt wurden. Zudem ist anzuführen, dass in den von Oblak et al. und Kida et al. aufgeführten Studien teilweise verschiedene Probandengruppen gegenübergestellt werden.[123]

Zur Bewertung der wissenschaftlichen Tragweite und des Impacts eines wissenschaftlichen Beitrages können bibliometrische Indikatoren wie die Anzahl der Zitationen und das Ranking des Journals, in welchem das jeweilige Paper veröffentlicht wurde, herangezogen werden.[124] Wie in Kapitel 2 beschrieben, werden durch die vorgenommene Eingrenzung des Suchumfangs lediglich Journale, welche ausgehend von dem VHB-Ranking A+- und A- geranked sind, untersucht und dargestellt. Daher können diese Beiträge als Teil der weltweit führenden und relevanten wissenschaftlichen Journale auf dem Gebiet der Betriebswirtschaftslehre im Bereich des Accountings angesehen werden. Nachfolgend beziehen sich die aufgeführten Zitationen auf die Datenbank Google Scholar.[125]

Im Bereich des Einsatzes neurowissenschaftlicher Methoden verzeichnen in der Kategorie Entscheidung Farrell et al. mit 41 Verweisen die höchste Zitationshäufigkeit. Auch das von Chen et al. 2016 erschienene Paper zeigt eine relativ hohe Zitationshäufigkeit von 18 auf. Abschließend kann in dieser Kategorie der Beitrag von Fehrenbacher et al. mit lediglich einer dokumentierten Zitation aufgeführt werden. In der Kategorie Verhaltenssteuerung sind mit meiner Zitationshäufigkeit von 232 seit Veröffentlichung im Jahr 1997 Hunton et al. anzuführen. Darauffolgend kann der Beitrag von Barton et al. mit einer Zitationshäufigkeit von 26 seit Veröffentlichung im Jahre 2014 angeführt werden. Abschließend kann in der Kategorie Verhaltenssteuerung der Beitrag von Escenazi et al. genannt werden, welcher eine Zitationshäufigkeit von 17 aufweist.

Im Bereich der Beiträge, die auf NWE verweisen, kann in der Kategorie Entscheidungen das Paper von Kida et al. an erster Stelle angeführt werden. Dieser Beitrag wurde seit Veröffentlichung 119 Mal zitiert. Abschließend ist Oblak et al. mit lediglich einer dokumentierten Zitation anzuführen. In der Kategorie Verhaltenssteuerung sind Jia et al. mit der höchsten Zitationshäufigkeit von 84 seit Veröffentlichung im Jahr 2014 anzuführen. Darauffolgend kann der Beitrag von Maas et al. mit einer Zitationshäufigkeit von 63 angeführt werden. Abschließend zeigt das von Hecht et al. 2012 erschienene Paper ebenfalls eine relativ hohe Zitationshäufigkeit von 27 auf.

Zusammenfassend ist somit zu erkennen, dass der Verweis auf neurowissenschaftliche Studienergebnisse und der Einsatz neurowissenschaftlicher Messmethoden gemessen an der Anzahl der Veröffentlichungen überschaubar ist. Darüber hinaus variieren die Entscheidungsszenarien und die untersuchten Probanden zwischen Accounting-Beitrag und neurowissenschaftlichem Beitrag. Ferner unterscheiden sich die Beiträge sowohl in der Quantität als auch der Qualität der Verweise (kontextspezifische Verweise vs. Kausalschlüsse). Lediglich in zwei Accounting-Beiträgen werden kontextspezifische Verweise vorgenommen. Dennoch ist ein emergierender Forschungsstrang des Neuroaccounting zu erkennen. Hierbei ist eine Zuordnung der identifizierten Beiträge zu den Kategorien Entscheidung und Verhaltenssteuerung möglich. In Bezug auf die von Birnberg et al. prognostizierten Phasen des Neuroaccounting ist erkennbar, dass ausgehend von dem Veröffentlichungsjahr erste Veröffentlichungen im Neuroaccounting zunächst auf neurowissenschaftliche Studienergebnisse verweisen und erst später selbst neurowissenschaftliche Methoden in der eingesetzt werden. Hierbei ist jedoch anzumerken, dass die jüngsten Veröffentlichungen größtenteils weiterhin auf neurowissenschaftliche Erkenntnisse verweisen. Abschließend kann subsumiert werden, dass die von Birnberg et al. prognostizierten Phasen des Neuroaccounting zu erkennen sind. Diese Phasen des Neuroaccounting lassen sich jedoch nicht trennscharf voneinander separieren, sondern gehen nebeneinander her.

4. Bewertung der Übertragbarkeit von neurowiss. Erkenntnissen auf einen Accounting-Kontext

Abschließend sollen die gefunden Ergebnisse hinsichtlich der vorgenommenen Übertragbarkeit von neurowissenschaftlichen Studienergebnissen auf einen Accounting-Kontext diskutiert werden. Hierbei liegt der Fokus primär auf dem Entscheidungskontext, der Probandenwahl sowie der im Rahmen der neurowissenschaftlichen Studie untersuchten Variable(n) (Kapitel 4.1/4.2/4.3).

Wie in Kapitel 3.2 bereits dargestellt, ist gemessen an der Anzahl der Veröffentlichungen, der Einsatz neurowissenschaftlicher Messmethoden in der Accounting-Forschung überschaubar. Als Grund hierfür kann das Kosten-Nutzen-Verhältnis des Einsatzes neurowissenschaftlicher Messme-

[119] Hierzu siehe Sanfey et al. (2003), De Quervain et al. (2004).
[120] Hierzu siehe Singer et al. (2006).
[121] Vgl. Bland et al. (2017), S. 1.
[122] Vgl. Sanfey et al. (2003), S. 1755; De Quervain et al. (2004), S. 1258.
[123] Aufgrund der kontextspezifischen Untersuchung werden bei der von Oblak et al. verwiesenen Studie von Treadway et al. Probanden mit schizophrener Erkrankung (krank) und Probanden ohne schizophrene Erkrankung (gesund) untersucht. Die von Kida et al. verwiesene Studie von Bechara et al. stellt zudem Probanden mit Präfrontalschäden (krank) und Probanden ohne Präfrontalschäden (gesund) in ihrer Untersuchung gegenüber.
[124] Vgl. Havemann (2016), S. 107-111.
[125] Vgl. Google Scholar (2018).

thoden angeführt werden.[126] Forscher im Accounting verfügen i.d.R. nicht über das notwendige Wissen zur Analyse und Interpretation der daraus resultierenden Daten und Ergebnisse. Die Einführung neurowissenschaftlicher Methoden im Accounting ist daher mit erheblichen intellektuellen Kosten verbunden, die entweder aus Schulungs- und Ausbildungsmaßnahmen oder der Zusammenarbeit mit Neurowissenschaftler in einem Forschungsteam resultieren.[127] Des Weiteren kann aufgrund des hohen Schulungsaufwands und der interdisziplinären Ausrichtung von Seiten der Forscher eventuell kein Interesse bestehen, in diese Richtung ausgebildet zu werden. Darüber hinaus sind neurowissenschaftliche Messmethoden in betriebswirtschaftlichen Disziplinen aufgrund des bisher nur überschaubaren Einsatzes begrenzt verfügbar und die eingesetzten Technologien häufig sehr teuer. Vor diesem Hintergrund können hohe finanzielle Kosten bei der Planung des Forschungsprojekts folgen, die zu berücksichtigen sind.

Zur Diskussion der Übertragbarkeit von Studienergebnissen auf einen anderen Kontext liegt der Fokus der experimentellen Forschung vor allem auf der externen Validität als Gütekriterium empirischer Untersuchungen. Die externe Validität bezeichnet die Generalisierbarkeit (Induktion) von empirisch gefundenen Untersuchungsergebnissen auf andere Kontextsituationen.[128] Folglich handelt es sich hierbei um ein Situationskriterium.[129] Dabei gilt: Je höher die externe Validität, desto höher die Repräsentativität der Ergebnisse. Nachfolgend werden vor allem die drei Ebenen Untersuchungskontext, untersuchte Stichprobe sowie Untersuchungsschwerpunkt betrachtet.[130]

4.1. Bewertung des Untersuchungskontextes

Wie in Kapitel 3.2 tabellarisch dargestellt, verweist keiner der untersuchten Beiträge auf neurowissenschaftliche Erkenntnisse, die in einem Accounting-Kontext gefunden wurden. Lediglich in den von Hecht et al. verwiesenen neurowissenschaftlichen Studien wird ein betriebswirtschaftliches Entscheidungsszenario zugeschrieben. Ferner gibt es keine identifizierten Accounting-Beiträge, bei welchen das gleiche Aufgabendesign wie in den verwiesenen neurowissenschaftlichen Studien besteht. Vor allem psychologische Forschungsdisziplinen, die sich mit der menschlichen Entscheidungsfindung befassen, zeigen auf, dass der Entscheidungskontext eine wichtige Determinante für die gefundenen Ergebnisse darstellt.[131] Ben-Akiva et al. führen an, dass eine Präferenz im Rahmen einer Entscheidung stets als Interaktion zwischen dem (kognitiven) Auswahlprozess selbst und dem Kontext der Entscheidung zu verstehen ist.[132] Des Weiteren führt Schmid an, dass das Verhalten in Entscheidungssituationen ebenfalls von der physischen, soziokulturellen und psychologischen Umgebung der Probanden beeinflusst wird.[133] Ein besseres Verständnis der Kontextabhängigkeit ist auch für die externe Validität von Untersuchungsergebnissen wichtig. Denn Menschen fällen zu keinem Zeitpunkt Entscheidungen, die vollständig unabhängig von Einflüssen und Restriktionen der bestehenden Umwelt sind.[134] Deswegen spricht man im Rahmen der Situationsrepräsentanz auch von Situationsvalidität.[135]

Wie schon in Kapitel 3.2 aufgeführt, liegt bei keiner der zitierten neurowissenschaftlichen Studien ein Accounting-Kontext vor. Auch vor dem Hintergrund des untersuchten Investitionsverhaltens bei Oblak et al. und Kida et al. wird nicht auf neurowissenschaftliche Studienerkenntnisse verwiesen, die in einem betrieblichen Entscheidungskontext ermittelt und untersucht wurden. In der Kognitionspsychologie werden hierzu Kontext-Effekte untersucht, welche bspw. den Einfluss von affektiven Einflüssen auf die Wahrnehmung eines Reizes beschreiben, die sich aus der Kontextsituation ergeben.[136] Affektive Einflüsse wie positive oder negative Emotionen auf die Entscheidungsfindung stoßen auch in der betriebswirtschaftlichen Forschung immer mehr auf Interesse. Standard-ökonomische Theorien gehen bei einer Investitionsentscheidung von einem analytischen Prozess aus. Bei diesem Prozess nutzen Einzelpersonen Opportunitätskosten zur Abwägung des daraus resultierenden Kosten-Nutzenverhältnisses vor dem Hintergrund der Nutzenmaximierung.[137] Neuere verhaltensökonomische Theorien deuten jedoch auf einen weiteren, hedonischen Prozess hin, der als „Pain of Paying“ bezeichnet wird.[138] Hierbei wird angenommen, dass jedes Mal, wenn eine Zahlung/Investition getätigt wird, ein psychologischer Schmerz empfunden wird. In diesem Zusammenhang ist neurowissenschaftlich erwiesen, dass bei privaten Kaufentscheidungen diejenigen Bereiche im Gehirn durch einen Reiz aktiviert und stimuliert werden, die mit physischen Schmerzen verbunden sind.[139] Je stärker dieser Schmerz ist, desto geringer ist die Vorfreude auf die Investition, welche in einer Beeinflussung der Investitionsentscheidung resultiert. Die funktionale Rolle des „Pain of Payings“ besteht folglich darin, Verzerrungen bei der Bewertung von Kosten und Nutzen zum Zeitpunkt der Investition entgegenzuwirken. Durch psychische Stimulation soll eine Überschreitung der eigenen Ausgaben verhindert werden.[140]

Vor dem Hintergrund des Entscheidungskontextes kommt die Frage auf, inwiefern sich das „Pain of Paying“ in Abhängigkeit zum jeweiligen Kontext verändert und letztlich das

[126] Vgl. Birnberg und Ganguly (2012), S. 7.
[127] Vgl. Birnberg und Ganguly (2012), S. 7.
[128] Vgl. Sarris und Reiß (2005), S. 41.
[129] Vgl. Windeler (2008), S. 253.
[130] Die Untersuchung der beiden Ebenen Untersuchungskontext und Stichprobe erfolgen in Anlehnung an die von Campbell und Cook (1979) vorgeschlagene Definition der externen Validität (S. 83). Diese beiden Ebenen werden um die von Meyer (1995) vorgeschlagene Ebene Untersuchungsschwerpunkt erweitert (S. 153).
[131] Vgl. Swait et al. (2002), S. 195.
[132] Vgl. Ben-Akiva et al. (2012), S. 454f.
[133] Vgl. Schmid (1981), S. 105.
[134] Vgl. Friedrichs und Nonnenmacher (2010), S. 469.
[135] Vgl. Döring und Bortz (2016), S.57.
[136] Vgl. Lubow et al. (1976), S. 38.
[137] Vgl. Prelec und Loewenstein (1998), S. 4.
[138] Vgl. Zellermayer (1996), S. 2.
[139] Vgl. Knutson et al. (2007), S. 147.
[140] Vgl. Prelec und Loewenstein (1998), S. 25.

Entscheidungsverhalten beeinflusst. Die Intensität des „Pain of Paying“ ist somit maßgeblich von zwei Faktoren abhängig: dem Investitionsmittel und dem Investitionskontext.[141] Inwiefern sich die Intensität des „Pain of Paying“ jedoch angesichts eines privaten oder eines Accounting-Kontextes unterscheidet, wird in der Literatur bisher nicht diskutiert. Lediglich Erkenntnisse aus der Verhaltensforschung von Konsumenten zeigen, dass bargeldlose Zahlungsmethoden zu weniger transparenten Transaktionsmustern führen und daraus eine Milderung des „Pain of Paying“ -Effekts resultiert.[142] Grund hierfür ist, dass für den Investor aufgrund des weniger transparenten Transaktionsmusters eine weniger aktive Präsenz der eigenen finanziellen Mittel und daraus resultierende Reduzierung der Vermögensgegenstände zu erkennen ist.[143] Dies stellt folglich eine Verzerrung der bestehenden Investitionssituation dar.

Der zweite Einflussfaktor, der sog. Investitionskontext, beruht zumeist auf kontextspezifischen Motiven, die sich aus der jeweiligen Investitionssituation ergeben. In diesem Zusammenhang sind vor allem die unterschiedlichen Motive zwischen privaten Investitionsentscheidungen und betrieblichen Entscheidungen im Rahmen eines Accounting-Kontextes zu berücksichtigen. Hierbei wird vorrangig auf die Frage abgezielt, inwiefern sich Entscheidungen unterscheiden, wenn mit eigenen finanziellen Mittel eine Investition getätigt wird oder bspw. im Rahmen einer Budgetausgabe in betrieblichen Entscheidungssituationen. Bei letzteren Investitionen hängen die Entscheidungen nicht unmittelbar von den Motiven und den damit einhergehenden Effekten einer privaten Kaufentscheidung ab. Ferner ist die Anwesenheit der eigenen finanziellen Mittel zumeist nicht ersichtlich. In Verbindung mit der Annahme der neuen Institutionenökonomik, dass unternehmerische Entscheidungen von opportunistischem Verhalten bestimmt werden[144], lässt sich schließen, dass sich die Intensität des „Pain of Paying“ -Effekts je nach ändernden Zielvorstellungen sowie unterschiedlicher Präsenz der Finanzmittel verändern kann. Wird folglich auf neurowissenschaftliche Erkenntnisse verwiesen, in welchen Investitionsentscheidungen nicht in einem Accounting-Kontext getroffen wurden, so können Verzerrungen der Entscheidungen aufgrund des Pain of-Paying Effekts auftreten. Dies ist bei der Übertragung von neurowissenschaftlichen Erkenntnissen auf einen Accounting-Kontext zu beachten, um eine ausreichend hohe Repräsentativität der übernommenen Ergebnisse zu gewährleisten.

Auch in dem von Farrell et al. aufgezeigten Zusammenhang zwischen dem Vergütungsdesign und der Investitionsentscheidung findet sich die oben aufgeführte Verbindung wieder. Farrell et al. verdeutlichen, dass im Vergleich zu Festverträgen leistungsbasierte Verträge den Anteil wirtschaftlich kostspieliger Entscheidungen verringern.[145] Hierbei führen sie zudem auf, dass nach Auswertung der im Rahmen einer fMRT gefundenen Ergebnisse zu erkennen ist, dass die Intensität affektiver Einflüsse je nach Vertragsdesign unterschiedlich stark wirken. Rückwirkend lässt sich hier schließen, dass die aus dem leistungsbasierten Vertrag resultierende Präsenz der privaten finanziellen Mittel eher auf ein Vorhandensein des „Pain of Paying“- Effekts verweist als bei Festverträgen. Somit wird das private Kosten-Nutzenverhältnis des Investors tangiert und das private Transaktionsmuster transparenter. Vor diesem Hintergrund lässt sich schließen, dass mit zunehmender Transparenz der eigenen finanziellen Mittel auch die Risikoaversion hinsichtlich unternehmerischer Entscheidungen steigt, sofern die finanzielle Entlohnung an die Investitionsentscheidung gekoppelt ist. Dieser Umstand ist ebenfalls bei der Übertragung von neurowissenschaftlichen Studienergebnissen zu berücksichtigen. Durch den Einsatz eines Accounting-Kontextes, wie bei Farrell et al. dargestellt, ist die neuronale Lokalisation eines relevanten „Pain of Paying“-Effekts möglich und kann somit bei der Interpretation und statistischen Auswertung der gewonnen Daten berücksichtigt werden.

Um eine Repräsentativität bei der Übertragung von Studienergebnissen auf einen anderen Kontext zu gewährleisten, sind ebenfalls die kontextspezifischen Motive des Untersuchungsdesigns zu berücksichtigen.[146] Auch im Rahmen der identifizierten Accounting-Beiträge sind motivationale Unterschiede im Untersuchungsdesign vorhanden, welche sich aus den kontextspezifischen Gegebenheiten der Untersuchung ergeben. Hierbei ist u.a. auffällig, dass die von Hecht et al. aufgeführte neurowissenschaftliche Studien von Pochon et al. in einem Kontext ohne monetären Anreiz durchgeführt wird. Hecht et al. hingegen verwenden einen Accounting-Kontext mit monetärem Anreiz. Der monetäre Anreiz ist hier nicht in Form einer Entlohnung für die Teilnahme an der Studie selbst zu verstehen, sondern im spezifischen Aufgabendesign zu finden. Das bedeutet, dass aufgrund des Untersuchungsdesigns entweder eine Gewinn- oder Verlustsituation möglich ist. Verhaltenspsychologische Erkenntnisse zeigen dabei, dass extrinsische Gewinn- und Verlustmotive, die sich aus dem jeweiligen Aufgabendesign ergeben, Einfluss auf die intrinsische Motivation bei der Verarbeitung von Informationen haben können.[147] Folglich können aufgrund des spezifischen Aufgabendesigns abweichende Untersuchungsergebnisse auftreten, die aus kontextspezifischen Gewichtungseffekten bei der Informationsverarbeitung resultieren können. Dieser Umstand ist ebenfalls bei der Übertragung neurowissenschaftliche Erkenntnisse auf einen Accounting-Kontext zu berücksichtigen.

Ein weiteres weit verbreitetes psychologisches Phänomen in der Verhaltensforschung ist der sog. „Framing“-Effekt. Dieser kennzeichnet, dass es bei Variation der Darstellungsweise von Informationen bei gleichbleibender inhaltlicher Logik zu unterschiedlichen Entscheidungen kommt. Der Begriff

[141] Vgl. Zellermayer (1996), S. 20f.
[142] Vgl. Raghubir und Srivastava (2008), S. 214.
[143] Vgl. Soman (2001), S. 470f.
[144] Vgl. Williamson (1993), S. 98.
[145] Vgl. Farrell et al. (2014), S. 2004.
[146] Vgl. Dholakia und Bagozzi (2002), S. 170.
[147] Vgl. Deci (1971), S. 109.

„Framing“ bezeichnet dabei i.d.R. den sprachlichen Rahmen, in den eine Information eingebunden wird.[148] Durch die Art und Weise der Darstellung können verschiedene Befindlichkeits- und Motivationssituationen beim Menschen hervorgerufen werden, die in abweichenden Entscheidungen resultieren. Des Weiteren kann der sog. Anker-Effekt angeführt werden. Dieser beschreibt die Tendenz, erste Eindrücke zur Bildung von Wahrnehmungen zu nutzen. Anker können entweder als unbewusste Suggestion wirken, welche die dazu passenden Assoziationen aktiviert, welche im Anschluss die Urteilsfindung beeinflussen. Andernfalls können Anker durch einen Ausgangspunkt oder Startwert für einen bewussten Gedankengang verantwortlich sein.[149] Erkenntnisse in der bestehenden Aufmerksamkeitsforschung zeigen auf, dass der Anker-Effekt in visuelle oder numerische Anker unterschieden werden kann.[150] Hierbei ist bekannt, dass der visuelle Anker den Entscheidungsprozess signifikant zu beeinflussen scheint, aber keine Ergebniswirkung hat. Der numerische Anker hingegen hat einen signifikanten Einfluss auf das Entscheidungsergebnis.[151] Der Entscheidungsprozess spiegelt dabei die Art und Weise wieder, wie Benutzer mit visuellen Messsystemen wie dem Eye-Tracking interagieren. Als Output ist die endgültige Entscheidung zu verstehen, die jeweils von den Teilnehmern am Ende des Entscheidungsprozesses getroffen werden. Anhand dieser Erkenntnisse wird deutlich, dass auch bei psychophysiologischen Verfahren wie dem Eye Tracking Anker-Effekte auftreten können. Die von Hunton et al., Chen et al. und Fehrenbacher et al. in einem Accounting-Kontext gefundenen Erkenntnisse können somit von Anker-Effekten frei sein. Besteht aufgrund des Aufgabendesigns jedoch ein „Anker“, ist dies durch die spezifische Technologie des Eye-Tracking feststellbar und kann bei der Interpretation und Bewertung der Daten berücksichtigt werden. In experimentellen Studien, in denen jedoch keine visuellen Analysetools eingesetzt werden, ist das feststellen von Anker-Effekten visuell faktisch unmöglich. Folglich wird also deutlich, dass „Framing“- und Ankereffekte aus verschiedenen Aufgaben- und Untersuchungsdesigns resultieren können.

Des Weiteren können motivationale Unterschiede aufgeführt werden. Hecht et al. führen in einer Nicht-Konkurrenzsituation ihre wissenschaftliche Studie durch, verweisen jedoch auf neurowissenschaftliche Studienerkenntnisse von Locke et al., welche Untersuchungen in einer Konkurrenzsituation durchführen. Hierbei können durch unterschiedliche Aufgabendesigns „Framing“-Effekte auftreten. Aufgrund der bestehenden Konkurrenz- oder Nicht-Konkurrenzsituation kann es zu verschiedenen motivationalen Treibern in der Entscheidungsfindung kommen. In Konkurrenzsituationen ist zudem die Problematik opportunistischen Verhaltens zu erkennen. Diese Problematik ist jedoch in einer Nicht-Konkurrenzsituationen weniger präsent und kann daher den Output eines kognitiven Prozesses, d.h. die Entscheidung selbst, weniger beeinflussen als in Konkurrenzsituationen. Grund hierfür sind zumeist unterschiedliche motivationale Antreiber. Forschungsergebnisse von Sun et al. zeigen dabei auf, dass verschiedene Kontexte Auswirkungen auf die emotionale und motivationale Empfindung haben können. Sie veranschaulichen, dass durch die Kombination verschiedener emotionaler oder motivationaler Kontexte die Informationsverarbeitung verschieden ist und folglich die Entscheidungsfindung beeinflusst werden kann.[152] Hecht et al. argumentieren auf Basis neurowissenschaftlicher Erkenntnisse, dass neben physiologischen Reaktionen auch weitere Teilanreize Antworten oder Entscheidungen beeinflussen.[153] Hierbei führen sie Erkenntnisse von Engelmann et al. auf, welche feststellen, dass eine der großen Herausforderungen bei der Untersuchung der neuronalen Effekte von Motivation darin besteht, unspezifische Effekte von den spezifischeren (d.h. kontext- bzw. aufgabenspezifischen) Effekten zu trennen.[154] Vor dem Hintergrund, dass Hecht et al. eine experimentelle Studie mit Accounting-Kontext (Nicht-Konkurrenzsituation) durchführen, aber auf Erkenntnisse zum Thema Motivation aus Nicht-Accounting-Kontexten (Konkurrenzsituation) verweisen, verdeutlicht die von Engelmann et al. geschilderte Problematik.

Hunton et al., Eskenazi et al. und Chen et al. umgehen Teile der Kontexteffekte dahingehend, dass sie professionelle Accounting-Teilnehmer im Rahmen der jeweiligen Studie untersuchen. Auch Fehrenbacher et al., Barton et al. und Farrell et al. führen ihre Studie mit Studenten durch, welche Accounting-Erfahrung aufgrund ihres Studienhintergrunds oder Praxiserfahrung vorweisen können. Dennoch haben diese Studien alle gemein, dass aufgrund des Einsatzes neurowissenschaftlicher Messmethoden ein Feldexperiment nur bedingt möglich ist. Sowohl Hunton et al., Chen et al. als auch Eskenazi et al., Farrell et al. und Barton et al. führen folglich ihre Untersuchungen in einem Laborexperiment durch. Auch bei Maas et al., Hecht et al., Kida et al. und Oblak et al. erfolgt die Untersuchung in Form einer experimentellen Laboruntersuchung. Diese Vorgehensweise weist jedoch im Gegensatz zu einem Feldexperiment eine geringere externe Validität auf.[155] Eine nahezu perfekte Kontrolle der Störvariablen ist jedoch zwangsläufig nur bei überwachten, unnatürlichen Untersuchungssituationen möglich.[156] Die „unnatürliche“ Gestaltung der Versuchsumgebung lässt es jedoch oft fraglich erscheinen, ob die erhaltenen Ergebnisse auch auf natürlichere Situationen generalisierbar sind.[157] Döring et al. führen hierzu an, dass eine Durchführung der Untersuchung in Form einer experimentellen Felduntersuchung bezüglich interner und externer Validität den anderen Kombinationen vorzuziehen ist. Demzufolge können somit

[148] Vgl. Tversky und Kahneman (1981), S. 453.
[149] Vgl. Tversky und Kahneman (1974), S. 1128.
[150] Vgl. Cho et al. (2017), S. 10.
[151] Vgl. Cho et al. (2017), S. 10.
[152] Vgl. Sun et al. (2018), S. 11.
[153] Vgl. Hecht et al. (2012), S. 567.
[154] Vgl. Engelmann et al. (2009), S. 12.
[155] Vgl. Renner et al. (2012), S. 74.
[156] Vgl. Hussy et al. (2012), S. 145.
[157] Vgl. Stroebe (1980), S.28; Döring und Bortz (2016), S. 57.

für hypothesenforschende Studien die eindeutigste Ergebnisinterpretation sowie Möglichkeit zur Generalisierung der Ergebnisse gewährleistet werden.[158] Aufgrund der eingesetzten Messmethoden und aufgabenspezifischen Problem- und Fragestellung, kann die Implementierung einer experimentellen Felduntersuchung zu einem nicht effizienten Kosten-Nutzenverhältnis einer wissenschaftlichen Studie führen. Daher wird häufig auf künstliche Untersuchungsbedingungen in Form einer Laboruntersuchung zurückgegriffen.[159] Zudem lassen, wie schon in Kapitel 2 aufgeführt, neurowissenschaftliche Messmethoden wie ein EEG oder fMRT aufgrund der großen und häufig tendenziell eher schlecht transportierbaren Messgeräte Felduntersuchungen nur bedingt zu.[160] Wie schon zu Beginn dieses Kapitels aufgeführt, ist häufig das bestehende Kosten-Nutzenverhältnis nicht effizient.

Abschließend ist zu der Ebene Kontext anzumerken, dass aufgrund des variierenden Aufgabendesigns und der daraus resultierenden Kontextunterschiede, psychologische Effekte wie der „Pain of Paying"- Effekt, der „Framing"- Effekt sowie der Anker-Effekt bei der Übertragbarkeit neurowissenschaftlicher Studienergebnisse auf einen Accounting-Kontext zu berücksichtigen sind. Zudem können aufgrund unterschiedlicher motivationaler Aufgabendesigns deutliche Unterschiede in der Ergebnisfindung auftreten. Swait et al. führen hierzu an, dass das nicht eindeutige Darstellen kontextspezifischer Variablen dazu führt, dass die Vorhersage eines gewissen Verhaltens oder die daraus resultierenden Entscheidung nur unter bestimmten Bedingungen korrekt sind. Dies ist nur dann der Fall, wenn der Kontext der realen Entscheidung ähnlich genug ist, um die Umstände und den Effekt der experimentellen Entscheidungssituation abzubilden.[161] Dies geht folglich mit der externen Validität der gefundenen Untersuchungsergebnisse einher.

4.2. Bewertung der Stichprobe

Die zweite Ebene der Übertragbarkeit von neurowissenschaftlichen Studienergebnissen auf einen Accounting-Kontext legt den Fokus auf die Stichprobe, die während eines experimentellen Designs untersucht wird. Hierbei ist zu berücksichtigen, dass eine Verallgemeinerung der Untersuchungsergebnisse nur dann aussagekräftig ist, wenn eine repräsentative Stichprobe untersucht wird. Repräsentativ ist eine Stichprobe, welche die Grundgesamtheit möglichst realistisch sowohl in quantitativer als auch qualitativer Ausführung abbildet. Deswegen spricht man im Rahmen der Stichprobenrepräsentanz auch von Populationsvalidität.[162] Vor dem Hintergrund der hier bestehenden Diskussion wird als Grundgesamtheit derjenige Personenkreis in einem Accounting-Kontext zu verstanden, der mit Accounting-spezifischen Aufgaben vertraut ist sowie selbst Accounting-Aufgaben bearbeitet.

Auffällig hierbei ist, dass Oblak et al. überwiegend auf Erkenntnisse aus Studien zurückgreifen, bei welchen nicht-humane Primaten im Rahmen experimenteller Studien untersucht werden. Vor diesem Hintergrund verweisen Oblak et al. auf Erkenntnisse, deren Stichprobe sich entweder aus Nagetieren[163] oder aus Affen und Nagetieren[164] zusammensetzt. Der Einsatz nicht-humaner Primaten zur Vorhersage von Aktionen und Reaktionen des Menschen, insbesondere in der Toxikologie und Pharmakologie, trifft auf Kontroversen in der Forschung hinsichtlich der Übertragbarkeit der Ergebnisse.[165] Inwiefern Ratten und andere nicht-humanen Primaten als Prädiktoren menschlichen Handelns in betrieblichen Entscheidungssituationen herangezogen werden können, ist in der Literatur bisher nicht abschließend beantwortet.

Ergebnisse von Steiner et al. zeigen bisher als einzige in der Forschung auf, dass Limitationen in der Informationsverarbeitung von nicht-humanen Primaten in einem neuroökonomischen-Kontext bestehen, die letztlich den Output, d.h. die Entscheidung selbst, beeinflussen. Hierbei steht primär das psychologische Phänomen des Bedauerns im Fokus. Steiner et al. zeigen dabei auf, dass bei Ratten sowohl der orbitofrontale Kortex als auch das ventrale Striatum[166] in die Entscheidungsfindung einbezogen wird. Dabei entdeckten Steiner et al., dass Ratten gelegentlich die bessere Alternative übersprangen und sich für die schlechtere Alternative entschieden. In diesem Zusammenhang konnten Steiner et al. beobachten, dass Ratten in diesen Situationen innehielten und anschließend nach einer kurzen Verzögerung wieder die bessere Alternative wählten.[167] Aus diesem Verhalten schließen Steiner et al., dass bedauernswerte Prozesse die Entscheidungsfindung bei Nagetieren verändern. Dies geht ebenfalls einher mit der in der Betriebswirtschaft bestehenden Theorie des Bedauerns, auch „Loss Aversion Bias" genannt. Bei einem „Loss Aversion Bias" wird davon ausgegangen, dass in ökonomischen Situationen aufgrund von Unsicherheit die Emotion des Bedauerns auftritt, welche die eigentliche Entscheidungsfindung in diesem Kontext beeinflusst.[168] Investoren, die von Bedauern beeinflusst werden, gehen ein geringeres Risiko ein, um das Auftreten schlechter Ergebnisse zu verringern. Der Prozess des Bedauerns wird somit als negative Emotion aufgefasst und ist von zentraler Bedeutung im Hinblick auf die bestehende Risikoaversion in Entscheidungssituationen. Inwiefern sich jedoch die Intensität dieses Bedauerns zwischen Menschen und nicht-humanen Primaten unterscheidet, wurde in der Forschung bisher noch nicht untersucht. Dies ist jedoch zur Gewährleistung der eindeutigen Übertragbarkeit von Verhalten nicht-humaner Primaten auf menschliches Verhalten zu berücksichtigen.

[158] Vgl. Döring und Bortz (2016), S. 57f.
[159] Vgl. Döring und Bortz (2016), S. 57.
[160] Vgl. Kenning (2014), S. 88.
[161] Vgl. Swait et al. (2002), S. 203.
[162] Vgl. Hussy et al. (2012), S. 118.
[163] Hierzu siehe Cocker et al. (2012); Hosking et al. (2014a); Hosking et al. (2014b) und Salamone et al. (1994).
[164] Hierzu siehe Walton et al. (2006).
[165] Vgl. Shanks et al. (2009), S. 4.
[166] Der orbitofrontale Kortex und das ventrale Striatum spielen eine wichtige Rolle bei der Entscheidungsfindung und bilden die Grundlage extensiver Forschungen menschlicher Entscheidungen (vgl. Schwarz (2018), S. 30.).
[167] Vgl. Steiner und Redish (2014), S. 995.
[168] Vgl. Zeelenberg (1999), S. 93.

Des Weiteren ist anzuführen, dass nicht-humane Primaten wie Ratten häufig aus der Motivation eines intuitiven Überlebensinstinktes heraus handeln.[169] Dies unterscheidet sich jedoch von den in betriebswirtschaftlichen Teildisziplinen bestehenden Motiven. Hierbei können vor allem Unterschiede sowohl in der bestehenden intrinsischen als auch extrinsischen Motivation zwischen den verschiedenen Stichprobenarten aufgeführt werden. Nicht-humane Primaten wie Nagetiere oder Affen werden mit Lebensmitteln angereizt. Dieser Anreiz resultiert letztlich in einer Reduzierung des Hungers und stellt einen motivationalen Treiber der Entscheidung dar.[170] Dieser Motivator ist u.a. auch bei der neurowissenschaftlichen Studie von Salamone et al. zu finden, auf die von Oblak et al verwiesen wird. Obwohl Nahrung und Essen als Notwendigkeit zur Reduzierung des Hungers auch bei Menschen nachgewiesen werden kann, ist Nahrung für die Mehrheit der Menschen in ökonomischen Entscheidungssituationen kein vorrangiger Motivator.[171] Bei betrieblichen Entscheidungssituationen in einem Accounting-Kontext treten i.d.R. abweichende extrinsische und intrinsische Motivatoren in den Vordergrund. Idealerweise werden hier Entscheidungsträger nicht von Überlebensinstinkten geleitet, sondern von dem Ziel der Unternehmensmaximierung oder auch finanziellen Anreizen in Form von leistungsabhängigen Entlohnungen oder Boni. Psychologische Forschungen von Hermalin et al. führen dabei auf, dass der aktuelle emotionale Zustand einen signifikanten Einfluss auf die Entscheidungsfindung, den Problemlösungsprozess sowie das Verhalten selbst hat.[172] Die aufgezeigten motivationalen Unterschiede sind ebenfalls bei der Übertragung von neurowissenschaftlichen Erkenntnissen auf einen Accounting-Kontext zu berücksichtigen.

Des Weiteren können kontextspezifische und untersuchungsrelevante Umstände wie Berufserfahrung und Bildungshintergrund in experimentellen Studien mit nicht-humanen Primaten nicht eindeutig dargestellt werden. Diese Umstände können lediglich mit Studienprobanden abgebildet werden, welche mit der Grundgesamtheit eindeutig übereinstimmen. Hierbei kann jedoch auch argumentiert werden, dass durch den Einsatz nicht-humaner Primaten die aus menschlichen Erfahrungs- und Wissensunterschieden resultierende Störvariablen, welche ebenfalls zu einer Verzerrung der Entscheidung führen[173], ausgeschlossen werden können. Hierbei ist zu berücksichtigen, dass der Mensch im Gegensatz zu Affen und Nagetieren komplexe Entscheidungsaufgaben mit minimaler Anleitung leicht erfüllen kann, wohingegen nicht-humane Primaten oft lange und intensive Eingewöhnungs- und Trainingsphasen benötigen, bis der eigentliche Versuch durchgeführt werden kann.[174] In diesem Zusammenhang kann es zu Lern- und Erfahrungseffekten kommen, die zu Verzerrungen des Entscheidungsprozesses führen können und bei der Bewertung der gefundenen Daten zu berücksichtigen sind.[175]

Zuletzt kann das Kosten-Nutzenverhältnis des Einsatzes nicht-humaner Primaten angeführt werden. Der Einsatz nicht-humaner Primaten resultiert überwiegend in geringeren Investitions- und Beschaffungskosten und bietet die Möglichkeit, entstandene Kosteneinsparungen in besseren Geräten oder umfangreicheren Studien und Tests im Rahmen der jeweiligen Forschungsarbeit einzusetzen.[176] Vor dem Hintergrund humaner Probanden bilden die von Jia et al. verwiesenen neurowissenschaftlichen Studien im Vergleich zu anderen Accounting-Beiträgen die untersuchte Grundgesamtheit (männliche CEOs) relativ genau ab. Grund hierfür ist, dass zumeist Männer untersucht werden oder Frauen ausreichende Mengen Testosteron verabreicht wurden, um messbare Unterschiede zu erkennen. Damit können Auswirkungen aus Stichproben-spezifischen Effekten grundsätzlich eingegrenzt werden.

Neben der geeigneten Probandenauswahl spielt die Stichprobengröße ebenfalls eine wichtige Rolle, um eine ausreichende Repräsentativität und anschließende Generalisierbarkeit experimenteller Ergebnisse zu gewährleisten. Eine Verallgemeinerung der Ergebnisse ist nur bei einer ausreichend großen Stichprobengröße mit hinreichender Genauigkeit auf die Grundgesamtheit gewährleistet. In neurowissenschaftlichen Untersuchungen werden aufgrund der sehr aufwendigen Technologien zumeist nur kleine Stichprobengrößen untersucht. Ausgehend davon wird im Anschluss auf eine Gesamtheit geschlossen. Dies zeigen auch die in Kapitel 3.2 aufgeführten Beiträge. Die von Hecht et al. verwiesene neurowissenschaftliche Studie von Pochon et al. wurde mit lediglich sechs Teilnehmern durchgeführt und auch auf die von Taylor et al. verwiesenen Erkenntnisse wurden mit 12 Teilnehmern gewonnen. Die von Kida et al. aufgeführte Studie von Bechara et al. weißt lediglich eine Stichprobengröße von 19 Teilnehmern auf. Ferner führen die von Maas et al. verwiesenen neurowissenschaftliche Studien von De Quervain et al. und Sanfey et al. eine Stichprobengröße von 15 bzw. 19 Untersuchungsprobanden an. In diesem Zusammenhang gilt jedoch, dass je größer die Stichprobe ist, desto höher ist auch die Wahrscheinlichkeit einer repräsentativen Messung. Durch diesen Umstand kann die Abbildung aller Merkmale präziser gewährleistet als bei einer kleineren Stichprobe.[177]

Im Allgemeinen kann angeführt werden, dass eine Stichprobengröße umso eher die Grundgesamtheit abdeckt, je eher sich die Größe der Stichprobe der Größe der Grundgesamtheit nähert. Zudem ist die Verallgemeinerung der Untersuchungsergebnisse nur dann aussagekräftig, wenn eine repräsentative Stichprobe untersucht wird, welche die Grundgesamtheit möglichst realistisch in quantitativer und qualitativer Ausführung abbildet.[178] Hierbei werden vor al-

[169] Vgl. Tuthill (2015), S. 1280.
[170] Vgl. Donovan (2015), S. 84.
[171] Vgl. Donovan (2015), S. 85.
[172] Vgl. Nooraie (2012), S. 405.
[173] Vgl. Ostrom et al. (1992), S. 300.
[174] Vgl. Hanks und Summerfield (2017), S. 20.
[175] Vgl. Westlund (2015), S. 119.
[176] Vgl. Hanks und Summerfield (2017), S. 18.
[177] Vgl. Hussy et al. (2012), S. 119.
[178] Vgl. Hussy et al. (2012), S. 118.

lem Erfahrung, persönlicher Hintergrund sowie weitere personenspezifische Merkmale angesprochen. Dennoch besteht im Rahmen einer geringeren Anzahl von Untersuchungsteilnehmern die Möglichkeit, eine größere Anzahl von Merkmalen umsichtiger und kontrollierter erfassen zu können.[179]

Unstimmigkeiten hinsichtlich unabhängiger Variablen wie Intelligenz und Inzentive der Teilnehmer werden in experimentellen Studien durch die Randomisierung der Teilnehmer minimiert. So können bei durchschnittlich hinreichend großen Gruppen relevante Störvariablen in allen Gruppen approximativ ausgeglichen werden.[180] In den Neurowissenschaften hingegen erfolgt häufig eine spezifische Zuteilung der Probanden, da überwiegend verschiedene Probandengruppen untersucht und miteinander verglichen werden. Auch die verwiesenen Erkenntnisse von Treadway et al., Bechara et al., Penton-Voak et al., Verplaetsea et al. und Zuckermann et al. untersuchen verschiedene Probandengruppen (gesund vs. „nicht"-gesund), welche spezifische Merkmale aufweisen. Hierbei wird bewusst ein Ausgleich wirkender gruppenspezifischer Variablen, die sich aus den verschiedenen Probandengruppen ergeben, verhindert, da diese hier die zu untersuchenden Variablen darstellen. Dieser Umstand ist bei der Übertragung von neurowissenschaftlichen Erkenntnissen auf einen Accounting-Kontext zu berücksichtigen.

4.3. Bewertung der untersuchten Variablen

Die letzte Ebene der externen Validität im Rahmen der Generalisierbarkeit von Untersuchungsergebnissen vor dem Hintergrund eines Accounting-Kontextes stellt die untersuchte Variable dar. Diese bildet die Grundlage für die Interpretation der gewonnen Daten und den logischen Kern der eigentlichen Untersuchung.[181] Die untersuchten Variablen werden generell im Rahmen der sog. Kriteriumsvalidität betrachtet. Von den untersuchten Variablen sollten sich Rückschlüsse auf die zugrundeliegenden thematischen Konstrukte ziehen lassen, welche nachfolgend als kausale Zusammenhänge verstanden werden.

Die von Jia et al. verwiesenen neurowissenschaftlichen Studien untersuchen ebenfalls die Auswirkungen von Testosteron auf verschiedene Attribute oder Verhaltensweisen. Die somit gezogenen Kausalschlüsse basieren auf der gleichen Untersuchungsbasis (Testosteron) und ermöglichen eine ausreichende Repräsentativität der Ergebnisse.

Oblak et al. hingegen beziehen sich bei der theoretischen Argumentation ihrer Ergebnisse auf neurowissenschaftliche Studien, deren Untersuchungsgegenstand abweichende Variablen darstellen. Wardle et al. untersuchen die Auswirkungen von Koffein auf die Bereitschaft, sich für monetäre Belohnungen einzusetzen. Hierbei wird der Zusammenhang zwischen Höhe des Koffeingehalts und der daraus resultierenden Aufwand untersucht.[182] Oblak et al. schließen daraus, dass Entscheidungen über Risiko und Aufwand miteinander verbunden sind, da sich die bei beiden Arten von Entscheidungen aktivierten neuronalen Netze tendenziell überschneiden.[183] Dass jedoch im Rahmen der eigentlichen Untersuchung Wadlaks et al. die untersuchte Variable die Beziehung zwischen Koffeingehalt und dem daraus resultierenden Aufwand untersucht und Entscheidungen über Risiko nicht betrachtet werden, geht aus der Argumentation Oblaks et al. nicht hervor. Zwar zeigen Wadlak et al. auf, dass unter der Bedingung verschiedener Koffeingehalte die neuronale Aktivität bei Entscheidungen über Aufwand gemessen werden. Inwiefern jedoch Entscheidungen über Risiko zu überlappender neuronaler Aktivität führen, wird im Rahmen dieser Untersuchung nicht betrachtet. Des Weiteren verweisen Oblak et al. auf neurowissenschaftliche Erkenntnisse von Cocker et al., deren Untersuchungsvariable ebenfalls nicht in unmittelbarem Zusammenhang mit den von Oblak et al. angeführten Zusammenhängen stehen. Auch hier werden keine Entscheidungen über Risiko untersucht, sondern lediglich aufgezeigt, dass Ratten bei der Entscheidungsfindung unterschiedlich empfindlich auf kognitive Anstrengungen reagieren und dabei verschiedene Gehirnareale stimuliert werden.[184]

Ferner untersuchen die Oblak et al. verwiesenen neurowissenschaftlichen Erkenntnisse von Hughes et al. und Treadway et al. grundsätzlich die Motivation der Probanden im Rahmen eines „Effort Expenditure for Rewards"- Aufgabendesigns. Dieses Aufgabendesign dient in der Psychologie in erster Linie der Untersuchung von Anhedonie[185] und aufwandsbasierten Entscheidungen bei schizophrenen Erkrankungen. Das Untersuchungsziel von Treadway et al. ist dabei, aufzuzeigen, wie die aufwandsbasierte Entscheidungsfindung mit dem Merkmal Anhedonie zusammenhängt. Hierbei wird deutlich, dass Oblak et al. nicht explizit auf die im Rahmen der jeweiligen neurowissenschaftlichen Studien betrachteten Variablen sondern lediglich im Rahmen der Untersuchung auf erkennbare Zusammenhänge verweisen und diese zur theoretischen Argumentation heranziehen. In diesem Zusammenhang ist jedoch zu berücksichtigen, dass die von Oblak et al. vorgenommene Verweislogik nicht auf die spezifischen Erkenntnisse der jeweiligen Studien abzielt sondern lediglich mögliche kausale Zusammenhänge aufzeigt, die aber in dem jeweiligen Untersuchungskontext der neurowissenschaftlichen Studie nicht Teil der untersuchten Variablen waren.

Maas et al. hingegen verweisen auf neurowissenschaftliche Erkenntnisse von Sanfey et al., Singer et al. und De Quervain et al., die primär mit den untersuchten Variablen der jeweiligen Studie zusammenhängen und gewährleisten somit eine logisch-schlüssige Kausalargumentation. Die von Sanfey et al. und Singer et al. untersuchte Korrelation zwischen Fairness bzw. Unfairness und neuronaler Aktivität, sowie die von

[179] Vgl. Döring und Bortz (2016), S. 395.
[180] Vgl. Fisher (1935), S. 23f; Huber (2009), S. 97; Döring und Bortz (2016), S. 54.
[181] Vgl. Meyer (1995), S. 153.
[182] Vgl. Wardle et al. (2012), S. 526.
[183] Vgl. Oblak et al. (2018), S. 35f.
[184] Vgl. Cocker et al. (2012), S. 1833.
[185] Anhedonie bezeichnet dabei die verminderte Fähigkeit zum Empfinden von Emotionen wie Freude oder Vergnügen und tritt häufig im Zusammenhang mit schizophrenen Erkrankungen auf (vgl. Horan et al. (2005), S. 259.).

De Quervain et al. untersuchten Korrelation zwischen Vertrauen und Bestrafung, bildet die Grundlage der kausalen Argumentation von Maas et al.. Da Maas et al. im Gegensatz zu Oblak et al. unmittelbar auf die genannten neurowissenschaftlichen Erkenntnisse verweist, ist ein logischer Kausalschluss gewährleistet. Daher sind in sich logische Kausalschlüsse vor dem Hintergrund der untersuchten Variablen bei der Übertragung neurowissenschaftlicher Ergebnisse ebenfalls zu berücksichtigen.

Die Problematik abweichender Untersuchungsvariablen umgehen Farrell et al. und Barton et al. dahingehend, da sie im Rahmen der formulierten Hypothesen die zu untersuchenden Variablen durch den Einsatz neurowissenschaftlicher Methoden direkt untersuchen und somit eine repräsentative Vorgehensweise gewährleisten. In diesem Zusammenhang ist jedoch auch die sog. prognostische Validität zu berücksichtigen, welche aufzeigt, wie gut von einem gemessenen Merkmal auf daraus resultierendes Verhalten außerhalb der eigentlichen Testsituation geschlossen werden kann. Die Enge dieser betrachteten Beziehung stellt das Ausmaß der Kriteriumsvalidität dar, welche ebenfalls als Korrelationsschluss bezeichnet wird.[186]

In neurowissenschaftlichen Erkenntnissen wird aufgrund der bildgebenden Methodiken überwiegend von physiologischen Prozessen auf daraus resultierendes Verhalten geschlossen. Hierbei wird zumeist die in Kapitel 2.2 angesprochene Problematik der „reverse inference“ praktiziert. Bei der „reverse inference“ wird nicht, wie von der Logik eines kausalen Zusammenhangs angenommen, vom Prozess auf das Signal geschlossen, sondern vom Signal auf den eigentlichen Prozess.[187] Sowohl Farrell et al. als auch Barton et al. schließen von den während der Untersuchung registrierten Signalen auf den eigentlichen neuronalen Prozess, validieren die Ergebnisse jedoch anschließend durch weitere statistische Auswertungen. Lediglich durch eine anschließende statistische Überprüfung der durch bildgebende Verfahren ermittelten Daten, ist eine logische Kausalität gewährleistet.[188] Bei der Übertragbarkeit neurowissenschaftlicher Studienergebnisse auf einen Accounting-Kontext ist somit ebenfalls die untersuchte Variable im Rahmen der eigentlichen Untersuchung zu berücksichtigen, um die Zulässigkeit und Gültigkeit der angewandten Kausalschlüsse zu gewährleisten.

Abschließend kann zur Bewertung der Übertragbarkeit von neuronalen Studienergebnissen auf einen Accounting-Kontext die Tragweite eines wissenschaftlichen Beitrages angeführt werden. Der Einfluss eines Papers auf zeitlich danach liegende Beiträge, gemessen an der Anzahl der Zitationen fällt umso stärker aus, je höher die Zitationshäufigkeit ist. Folglich ist diese ein Indikator dafür, inwieweit die gefundenen Erkenntnisse Anklang in nachfolgenden Forschungsarbeiten finden und wie neu und einzigartig die Ergebnisse selbst sind. Dies kann anhand der der in Kapitel 3.2 angeführten bibliometrischen Indikatoren geprüft werden. In diesem Zusammenhang ist jedoch anzuführen, dass bibliometrische Indikatoren auf Basis von Autoren aufgrund geringer Publikationszahlen nur bedingt aussagekräftig sind.[189] Zudem lassen sich infolge der Aktualität der Beiträge von Oblak et al. und Fehrenbacher et al. nur bedingt Aussagen über die tatsächliche wissenschaftliche Reichweite der Beiträge treffen. Vor diesem Hintergrund ist eine repräsentative und aussagekräftige Bewertung hinsichtlich der Tragweite im Vergleich zu anderen Beiträgen nur bedingt möglich. Des Weiteren dient als Datenbasis für die Anzahl der Zitationen nur die jeweilige betrachtete Datenbank. Infolgedessen ist eine exakte Anzahl aller Zitationen zu einer Publikation nicht darstellbar ist, da keine Datenbank alle Publikationen enthält.[190]

Ausgehend von der Anzahl der Zitationen ist erkennbar, dass vor allem Erkenntnisse derjenigen Beiträge, welche neurowissenschaftliche Messmethoden im Kontext Accounting-spezifischer Problem- und Fragestellungen einsetzen, in der nachfolgenden Forschung stark an Bedeutung gewinnen. Grund hierfür kann vor allem in der bisher grundlegend einzigartigen Identifikation neuronaler Vorgehensweisen im direkten Zusammenhang mit dem Accounting sein. Auch die vorher dargestellten Beiträge, welche auf bereits gefundene neurowissenschaftliche Erkenntnisse verweisen, finden vermehrt Anwendung in nachfolgenden Forschungsdisziplinen. Es ist jedoch zu berücksichtigen, dass diese nicht aufgrund des Verweises auf neurowissenschaftliche Erkenntnisse selbst in nachfolgenden Forschungsarbeiten wiederzufinden sind, sondern vielmehr aufgrund der im Rahmen gewöhnlicher experimenteller Forschungsmethoden gefundenen Erkenntnisse, welche letztlich nicht in unmittelbarem Zusammenhang mit dem eigentlichen Verweis stehen. Ein direkter Verweis darauf, dass zur theoretischen Argumentation ebenfalls neurowissenschaftliche Erkenntnisse herangezogen werden, wird in nachfolgenden Forschungsarbeiten bis dato nicht aufgeführt.

5. Konklusion

Nachfolgend werden die im Rahmen dieser Arbeit gefundenen Ergebnisse und die erläuterte Diskussion zusammengefasst. Im Anschluss ist ein resümierendes Fazit zu finden. Das Kapitel schließt mit einem kurzen Ausblick, welcher Implikationen für zukünftige Forschungen aufzeigt.

5.1. Zusammenfassung und Fazit

Der Forschungsstrang des Neuroaccountings kann in Anlehnung an Birnberg et al. in zwei Phasen unterteilt werden: Beiträge, die auf neurowissenschaftliche Erkenntnisse verweisen und Beiträge, die selbst neurowissenschaftliche Methoden nutzen, um Accounting-spezifische Problem- und Fragestellungen zu untersuchen. Aufgrund der gewonnenen Ergebnisse in Kapitel 3 ist jedoch anzumerken, dass diese

[186] Vgl. Moosbrugger und Kelava (2012), S. 18.
[187] Vgl. Bourgeois-Gironde (2010), S. 232.
[188] Vgl. Poldrack (2006), S. 59f.
[189] Vgl. Holsapple und Joshi (2004), S. 105.
[190] Vgl. Döring und Bortz (2016), S. 749.

Phasen nicht sequentiell, sondern viel mehr versetzt parallel laufen. In diesem Zusammenhang ist zu erkennen, dass vor allem Erkenntnisse derjenigen Beiträge, welche neurowissenschaftliche Messmethoden im Kontext Accounting-spezifischer Problem- und Fragestellungen einsetzen, in der nachfolgenden Forschung stark an Bedeutung finden. Grund hierfür kann vor allem in der bisher grundlegend einzigartigen Identifikation neuronaler Vorgehensweisen im direkten Zusammenhang mit dem Accounting stehen. Auch die aufgeführten Beiträge, welche auf neurowissenschaftliche Erkenntnisse verweisen, finden vermehrt Anwendung in nachfolgenden Forschungsdisziplinen. Hierbei ist jedoch zu berücksichtigen, dass diese nicht aufgrund des Verweises auf neurowissenschaftliche Erkenntnisse selbst in nachfolgenden Forschungsarbeiten wiederzufinden sind. Wird jedoch im Rahmen eines Accounting-Beitrages auf neurowissenschaftliche Erkenntnisse verwiesen, so unterscheiden sich die identifizierten Accounting-Beiträge sowohl in Qualität als auch Quantität der Verweise. Des Weiteren variieren die Stichproben der unterschiedlichen Beiträge stark voneinander. Hierbei ist auffällig, dass teilweise Erkenntnisse nicht-humaner Primaten herangezogen werden und überwiegend lediglich sehr kleine Probandengruppen untersucht werden. Der Verweis auf neurowissenschaftliche Erkenntnisse ermöglicht eine theoretisch fundierte Argumentation, um auch bei experimentellen Studien ohne den Einsatz neurowissenschaftlicher Messmethoden einen Blick in die „Black Box“ des Gehirns zu ermöglichen.

Vor dem Hintergrund, dass sich auch die Accounting-Forschung zunehmend für den Unterschied zwischen kognitiver (rationaler) und affektiver (emotionaler) Entscheidungsfindung interessiert, bietet eine Verknüpfung von Accounting-spezifischen Problemstellungen mit neurowissenschaftlichen Erkenntnissen neue Möglichkeit. So können neuronale Entscheidungszusammenhänge über Anreize, Risikobereitschaft, Umgang mit Unsicherheit, Motivation, Fairness und anderen Verhaltensergebnissen theoretisch fundierter untersucht werden. Hierbei sind jedoch vor allem Kriterien der externen Validität zu berücksichtigen, damit logische und repräsentative Kausalschlüsse gewährleistet werden können. Vor diesem Hintergrund ist zu beachten, dass interne und externe Validität nicht zwangsläufig konträr zueinanderstehen. Folglich ist ein Laborexperiment nicht zwingend aufgrund einer höheren internen mit einer niedrigeren externen Validität gekennzeichnet. Es handelt sich um situationsspezifische Gütekriterien, die je nach Studiendesign und eigentlichem Untersuchungskontext zu berücksichtigen sind. Infolgedessen kann ein Laborexperiment zugleich eine hohe interne als auch eine hohe externe Validität aufweisen.

Vor allem stichprobenrelevante Störvariablen sowie kontextspezifische psychologische Effekte können zu Limitationen bei der Übertragbarkeit führen. Folglich ist also eine generalisierende Bewertung der gefundenen Accounting-Beiträge, welche auf neurowissenschaftliche Erkenntnisse verweisen, nicht möglich. Die gefundenen Ergebnisse sind situations- und kontextabhängig. In diesem Zusammenhang sollte zudem beachtet werden, inwiefern im Kontext der jeweiligen Untersuchung die Verwendung neurowissenschaftlicher Erkenntnisse oder Methoden nützlich und notwendig ist, um relevante Ergebnisse für die das Accounting zu generieren.

In etablierten Theorien des Accounting werden bestehende ökonomische Problematiken wie affektive Einflüsse oder psychologische Effekte nur bedingt berücksichtigt. In diesen Theorien wird der menschliche Verstand, der letztlich der Treiber des wirtschaftlichen Handelns im Accounting ist, als “ Black Box“ betrachtet, die in Form von abstrakten Wahrscheinlichkeiten konzipiert ist, um die individuelle Handlung in einer gegebenen Situation zu beschreiben. Hierbei können durch den geeigneten Verweis auf neurowissenschaftliche Erkenntnisse oder den Einsatz neurowissenschaftlicher Messmethoden zuverlässige Forschungsergebnisse herangezogen und generiert werden. Folglich kann aufgezeigt werden, welche Verbindung zwischen affektiven und kognitiven Systemen bei der Entscheidungsfindung im Accounting bestehen und wie diese letztlich von (einer) affektiven Komponente(n) beeinflusst werden kann. Somit kann eine transparentere Darstellung der relevanten psychologischen Mechanismen innerhalb der menschlichen „Black Box“ gewährleistet werden. Letztlich können die zumeist auf statistisch- theoretischen Methoden basierenden Erkenntnisse des Accounting um eine mikroskopische Sichtweise erweitert werden. Zudem kann eine realitätsnähere Abbildung menschlichen Verhaltens gewährleistet werden. Vor diesem Hintergrund können Erkenntnisse des Neuroaccounting dazu beitragen, die Entscheidungs- und Verhaltenssteuerungsfunktion im Accounting auszubauen.

5.2. Implikationen für zukünftige neurowissenschaftliche Forschungen im Accounting

Der geringe Einsatz neurowissenschaftlicher Messmethoden im Accounting lässt sich zum einen auf die hohen Projektkosten, die aus dem Einsatz neurowissenschaftlicher Messmethoden im Accounting resultieren, zurückführen. Zum anderen können die nicht ausreichende fachliche Qualifikation zur Auswertung gefundener Daten sowie auf die begrenzte Verfügbarkeit dieser Messmethoden in betriebswirtschaftlichen Teildisziplinen aufgeführt werden. Das daraus resultierende Kosten-Nutzenverhältnis ist i.d.R. nicht effizient.

In Kombination mit dem wachsenden Interesse von Seiten der Forschung sowie der zunehmenden Anzahl von Veröffentlichungen im Neuroaccounting, kann angenommen werden, dass das Forschungsfeld des Neuroaccounting in der Zukunft vermehrt an Bedeutung gewinnen wird.

Um zu gewährleisten, dass aufgrund kontextspezifischer Störvariablen keine Limitationen bei der Übertragung von neurowissenschaftlichen Erkenntnissen auf einen Accounting-Kontext bestehen, müssen sowohl kontextspezifische Gegebenheiten, Stichprobeninterdependenzen als auch der eigentliche Untersuchungsschwerpunkt der neurowissenschaftlichen Studie berücksichtigt werden. Dabei können in zukünftigen Forschungen vor allem soziökonomische

und psychologische Verhaltenseffekte wie der „Pain of Paying“ -Effekt in Accounting-relevanten Untersuchungen durch den Einsatz neurowissenschaftlicher Messmethoden untersucht werden. Somit kann aufgezeigt werden, inwiefern Teile der bestehenden Unterschiede in Entscheidungssituationen durch das psychologische Phänomen des „Pain of Paying“ erklärt werden können. Des Weiteren können die in Kapitel 4 aufgezeigten Limitationspotenziale, hinsichtlich der untersuchten Stichprobe und den gewählten Entscheidungsszenarien bei künftigen Verweisen von neurowissenschaftlichen Studienerkenntnissen auf einen Accounting-Kontext, berücksichtigt werden. Vor diesem Hintergrund kann in künftigen Forschungen vor allem die Eignung nicht-humaner Primaten als Prädikatoren menschlichen Verhaltens in ökonomischen Situationen untersucht werden. Mit diesem Wissen kann die Qualität der Verweise auf neurowissenschaftliche Erkenntnisse deutlich verbessert werden.

Des Weiteren kann bei der bestehenden Verweislogik mehr auf spezifische neurowissenschaftliche Erkenntnisse verwiesen werden und weniger lediglich von einer Vielzahl von Beiträgen auf oberflächliche Erkenntnisse geschlossen werden. Ferner können bestehende Accounting-relevante Theorien und Erkenntnisse durch den Verweis auf neurowissenschaftliche Studienerkenntnisse oder den Einsatz neurowissenschaftlicher Methoden theoretisch fundierter dargestellt werden. Hierzu können bspw. die von Oblak et al. und Hecht et al. aufgeführten Erkenntnisse durch den Einsatz neurowissenschaftlicher Methoden mikroskopisch überprüft werden. Darüber hinaus können bestehende Theorien im Accounting durch neurowissenschaftliche Erkenntnisse validiert werden. Dadurch ergibt sich die Möglichkeit, realitätsferne Annahmen des Homo Oeconomicus zu widerlegen und somit eine realitätsnähere Abbildung menschlichen Verhaltens zu gewährleisten.

Um jedoch eine ausreichend starke Repräsentativität von Accounting-spezifischen Problem- und Fragestellung mit neurowissenschaftlichen Erkenntnissen zu gewährleisten, sind vor allem stichprobenrelevante Störvariablen in experimentellen Entscheidungssituationen zu eliminieren. In letzter Konsequenz ist dies jedoch erst dann ausreichend möglich, sofern der Einsatz neurowissenschaftlicher Methoden im jeweiligen Accounting-Kontext erfolgt. Hierzu können die im Rahmen dieser Arbeit identifizierten Limitationen in weiteren Forschungen untersucht werden.

Abschließend kann angeführt werden, dass stets Motive der interdisziplinären Zusammenarbeit und Bereitschaft für multidisziplinäre Forschungsvorgehensweisen gefördert werden müssen, um statisch hochwertige Ergebnisse durch den Einsatz neurowissenschaftlicher Methoden in experimentellen Studien im Accounting zu gewährleisten. Ferner können durch eine repräsentative Vorgehensweise bei der Verwendung von neurowissenschaftlichen Erkenntnissen und Methoden zumeist makrospezifische Erkenntnisse des Accountings durch mikrospezifische Erkenntnisse der Neurowissenschaften erweitert werden. Zudem kann die Unternehmensrechnung stärker auf das Verhalten von Entscheidungsträgern ausgereichtet werden.

Vor dem Hintergrund der in Kapitel 2.2 aufgeführten Vorteile neurowissenschaftlicher Methoden, kann vor allem auch die Art und Weise der Darstellung von Informationen untersucht werden. Somit kann das tatschliche Verhalten der Entscheidungsträger besser in die Unternehmensrechnung miteinbezogen werden. Resultierend daraus kann ein höherer Grad an Realitätsnähe der Ergebnisse erreicht werden. Abschließend können durch eine repräsentative Übertragung von neuronalen Studienergebnissen auf einen Accounting-Kontext vereinfachte theoretische Prämissen der normativen Entscheidungstheorie ergänzt und wissenschaftlich stärker fundiert werden. Dies kann sowohl in Form von repräsentativen neurowissenschaftlichen Verweisen oder auch den spezifischen Einsatz neurowissenschaftlicher Methoden in einem Accounting-Kontext erzielt werden.

Literatur

Amshoff, B. Controlling in deutschen Unternehmungen, 2. Aufl., Wiesbaden, Springer Gabler, 1993.

Apicella, C. L., Dreber, A., und Mollerstrom, J. Salivary testosterone change following monetary wins and losses predicts future financial risk-taking. *Psychoneuroendocrinology*, 39:58–64, 2014.

Ariely, D. und Berns, G. S. Neuromarketing: the hope and hype of neuroimaging in business. *Nature Reviews Neuroscience*, 11(4):284, 2010.

Artienwicz, N. Behavioral Stream in Polish Accounting, in: Christiansen, Bryan/Lechman, Ewa (Hrsg), Neuroeconomics and the decision-making process, 2016, Hershey, 9. Aufl.: 246-261, 2016.

Baker, M. J. Writing a literature review. *The Marketing Review*, 1(2):219–247, 2000.

Balthazard, P. A. und Thatcher, R. W. Neuroimaging modalities and brain technologies in the context of organizational neuroscience. In *Organizational Neuroscience*. Emerald Group Publishing Limited, Bingley: 83-113, 2015.

Barnes Jr, J. H. Cognitive biases and their impact on strategic planning. *Strategic Management Journal*, 5(2):129–137, 1984.

Barton, J., Berns, G. S., und Brooks, A. M. The neuroscience behind the stock market's reaction to corporate earnings news. *The Accounting Review*, 89 (6):1945–1977, 2014.

Bazire, M. und Brézillon, P. Understanding context before using it. In *International and Interdisciplinary Conference on Modeling and Using Context, Berlin, Springer: 29-40*, 2005.

Bechara, A., Tranel, D., Damasio, H., und Damasio, A. R. Failure to respond autonomically to anticipated future outcomes following damage to prefrontal cortex. *Cerebral Cortex*, 6(2):215–225, 1996.

Beck, S. M., Locke, H. S., Savine, A. C., Jimura, K., und Braver, T. S. Primary and secondary rewards differentially modulate neural activity dynamics during working memory. *PloS one*, 5(2):1–13, 2010.

Ben-Akiva, M., de Palma, A., McFadden, D., Abou-Zeid, M., Chiappori, P.-A., de Lapparent, M., Durlauf, S. N., Fosgerau, M., Fukuda, D., Hess, S., et al. Process and context in choice models. *Marketing Letters*, 23(2):439–456, 2012.

Beugré, C. D. *The Neuroscience of Organizational Behavior*. Cheltenham, Edward Elgar Publishing, 2018.

Birnberg, J. G. und Ganguly, A. R. Is neuroaccounting waiting in the wings? an essay. *Accounting, Organizations and Society*, 37(1):1–13, 2012.

Birnberg, J. G., Luft, J., und Shields, M. D. Psychology theory in management accounting research. *Handbooks of Management Accounting Research*, 1: 113–135, 2007.

Bland, A. R., Roiser, J. P., Mehta, M. A., Schei, T., Sahakian, B. J., Robbins, T. W., und Elliott, R. Cooperative behavior in the ultimatum game and prisoner's dilemma depends on players' contributions. *Frontiers in Psychology*, 8:1–11, 2017.

Blay, A., Douthit, J., und Fulmer III, B. Why don't people lie? negative affect intensity and preferences for honesty in budgetary reporting. *Management Accounting Research*, 42:56–65, 2019.

Bourgeois-Gironde, S. Is neuroeconomics doomed by the reverse inference fallacy? *Mind & Society*, 9(2):229–249, 2010.

Braeutigam, S. Neuroeconomics—from neural systems to economic behaviour. *Brain research bulletin*, 67(5):355–360, 2005.

Brammer, M. The role of neuroimaging in diagnosis and personalized medicine-current position and likely future directions. *Dialogues in Clinical Neuroscience*, 11(4):389–396, 2009.

Burgess, P. W. und Simons, J. S. Theories of frontal lobe executive function: clinical applications. *The Effectiveness of Rehabilitation for Cognitive Deficits*, pages 211–231, 2004.

Butler, M. J., O'Broin, H. L., Lee, N., und Senior, C. How organizational cognitive neuroscience can deepen understanding of managerial decision-making: A review of the recent literature and future directions. *International Journal of Management Reviews*, 18(4):542–559, 2016.

Cacioppo, J. T., Amaral, D. G., Blanchard, J. J., Cameron, J. L., Carter, C. S., Crews, D., Fiske, S., Heatherton, T., Johnson, M. K., Kozak, M. J., et al. Social neuroscience: Progress and implications for mental health. *Perspectives on Psychological Science*, 2(2):99–123, 2007.

Camerer, C., Loewenstein, G., und Prelec, D. Neuroeconomics: How neuroscience can inform economics. *Journal of Economic Literature*, 43(1): 9–64, 2005.

Camerer, C. F. Neuroeconomics: opening the gray box. *Neuron*, 60(3):416–419, 2008.

Camerer, C. F., Loewenstein, G., und Prelec, D. Neuroeconomics: Why economics needs brains. *Scandinavian Journal of Economics*, 106(3):555–579, 2004.

Camerer, C. F., Bhatt, M., und Hsu, M. Neuroeconomics: Illustrated by the study of ambiguity aversion. *Economics and Psychology, Cambridge*, pages 113–151, 2007.

Campbell, D. T. und Cook, T. D. *Quasi-experimentation: Design & analysis issues for field settings*. Chicago, Rand McNally College Publishing Company, 1979.

Caplan, E. H. Behavioral assumptions of management accounting. *The Accounting Review*, 41(3):496–509, 1966.

Carré, J. M., Gilchrist, J. D., Morrissey, M. D., und McCormick, C. M. Motivational and situational factors and the relationship between testosterone dynamics and human aggression during competition. *Biological Psychology*, 84(2):346–353, 2010.

Chen, Y., Jermias, J., und Panggabean, T. The role of visual attention in the managerial judgment of balanced-scorecard performance evaluation: Insights from using an eye-tracking device. *Journal of Accounting Research*, 54(1):113–146, 2016.

Chen, Z. J., He, Y., Rosa-Neto, P., Germann, J., und Evans, A. C. Revealing modular architecture of human brain structural networks by using cortical thickness from mri. *Cerebral Cortex*, 18(10):2374–2381, 2008.

Cho, I., Wesslen, R., Karduni, A., Santhanam, S., Shaikh, S., und Dou, W. The anchoring effect in decision-making with visual analytics. In *Phoenix, IEEE Conference 2017 on Visual Analytics Science and Technology (VAST): 116–126*, 2017.

Cocker, P. J., Hosking, J. G., Benoit, J., und Winstanley, C. A. Sensitivity to cognitive effort mediates psychostimulant effects on a novel rodent cost/benefit decision-making task. *Neuropsychopharmacology*, 37 (8):1825, 2012.

Cooper, H. M. Organizing knowledge syntheses: A taxonomy of literature reviews. *Knowledge in Society*, 1(1):104–126, 1988.

David, S. und Aguayo, A. J. Axonal elongation into peripheral nervous system 'bridges' after central nervous system injury in adult rats. *Science*, 214(4523):931–933, 1981.

De Quervain, D. J., Fischbacher, U., Treyer, V., Schellhammer, M., et al. The neural basis of altruistic punishment. *Science*, 305(5688):1254–1258, 2004.

Deci, E. L. Effects of externally mediated rewards on intrinsic motivation. *Journal of personality and Social Psychology*, 18(1):105–115, 1971.

Demski, J. S. und Feltham, G. A. *Cost determination: A conceptual approach*. Ames, Iowa State Pr, 1976.

Dholakia, U. M. und Bagozzi, R. P. Mustering motivation to enact decisions: How decision process characteristics influence goal realization. *Journal of Behavioral Decision Making*, 15(3):167–188, 2002.

Dickhaut, J., Basu, S., McCabe, K., und Waymire, G. Neuroaccounting: Consilience between the biologically evolved brain and culturally evolved accounting principles. *Accounting Horizons*, 24(2):221–255, 2010.

Donovan, N. The effects of intrinsic and extrinsic motivation on cognitive performance in humans. *The Huron University College Journal of Learning and Motivation*, 53(1):84–96, 2015.

Döring, N. und Bortz, J. Forschungsmethoden und evaluation. *Wiesbaden: Springerverlag*, 2016.

Drury, C. *Management and cost accounting*. 8. Aufl., Andover, 2012.

Duchowski, A. *Eye tracking methodology: Theory and practice*. London, Springer, 2017.

Dulleck, U., Ristl, A., Schaffner, M., und Torgler, B. Heart rate variability, the autonomic nervous system, and neuroeconomic experiments. *Journal of Neuroscience, Psychology, and Economics*, 4(2):117–124, 2011.

Edmunds, A. und Morris, A. The problem of information overload in business organisations: a review of the literature. *International Journal of Information Management*, 20(1):17–28, 2000.

Eisele, W. und Knobloch, A. P. *Technik des betrieblichen Rechnungswesens*. 8. Aufl., München, Vahlen, 2014.

Eisenegger, C., Naef, M., Snozzi, R., Heinrichs, M., und Fehr, E. Prejudice and truth about the effect of testosterone on human bargaining behaviour. *Nature*, 463(7279):356–359, 2010.

Engelmann, J. B., Damaraju, E., Padmala, S., und Pessoa, L. Combined effects of attention and motivation on visual task performance: transient

and sustained motivational effects. *Frontiers in Human Neuroscience*, 3: 1–17, 2009.

Eskenazi, P. I., Hartmann, F. G., und Rietdijk, W. J. Why controllers compromise on their fiduciary duties: Eeg evidence on the role of the human mirror neuron system. *Accounting, Organizations and Society*, 50:41–50, 2016.

Ewert, R. und Wagenhofer, A. *Interne Unternehmensrechnung*. Berlin, Springer, 2008.

Farrell, A. M., Goh, J. O., und White, B. J. The effect of performance-based incentive contracts on system 1 and system 2 processing in affective decision contexts: fmri and behavioral evidence. *The Accounting Review*, 89 (6):1979–2010, 2014.

Fehr, E., Fischbacher, U., und Kosfeld, M. Neuroeconomic foundations of trust and social preferences: initial evidence. *American Economic Review*, 95(2):346–351, 2005.

Fehrenbacher, D. D., Schulz, A. K.-D., und Rotaru, K. The moderating role of decision mode in subjective performance evaluation. *Management Accounting Research*, 41:1–10, 2018.

Fisher, R. A. The design of experiments. *Oxford, Oliver and Boyd*, 1935.

Fiske, S. T. und Taylor, S. E. *Social cognition: From brains to culture*. 2. Aufl., New York, Sage, 2013.

Friedrichs, J. und Nonnenmacher, A. Welche Mechanismen erklären Kontexteffekte? In *Komparative empirische Sozialforschung*. Wiesbaden, Springer: 469-497, 2010.

Gash, S. *Effective literature searching for research*. Aldershot, Gower, 2000.

Geniole, S. N., Keyes, A. E., Mondloch, C. J., Carré, J. M., und McCormick, C. M. Facing aggression: Cues differ for female versus male faces. *PLOS one*, 7(1):1–12, 2012.

Gillenkirch, R. M. State of the art des behavioral accounting. *WiSt-Wirtschaftswissenschaftliches Studium*, 37(3):128–134, 2008.

Glimcher, P. W. *Decisions, uncertainty, and the brain: The science of neuroeconomics*. Cambridge, MIT Press, 2004.

Google Scholar, 2018. URL https://scholar.google.de/scholar?cites=5737752200460274206&as_sdt=2005&sciodt=0,5&hl=de. [12.12.2018].

Haley, U. C. und Stumpf, S. A. Cognitive trails in strategic decision-making: linking theories of personalities and cognitions. *Journal of Management Studies*, 26(5):477–497, 1989.

Hanks, T. D. und Summerfield, C. Perceptual decision making in rodents, monkeys, and humans. *Neuron*, 93(1):15–31, 2017.

Havemann, F. *Einführung in die Bibliometrie*. Gesellschaft für Wissenschaftsforschung, Berlin, 2016.

Hecht, G., Tafkov, I., und Towry, K. L. Performance spillover in a multitask environment. *Contemporary Accounting Research*, 29(2):563–589, 2012.

Hofmann, C. und Küpper, H.-U. Neurobiologie und unternehmensrechnung. *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung*, 63(63): 168–196, 2011.

Holsapple, C. W. und Joshi, K. D. A knowledge management ontology. In *Handbook on Knowledge Management 1*. Holsapple, Clyde W., Heidelberg, Springer: 89-124, 2004.

Horan, W. P., Kring, A. M., und Blanchard, J. J. Anhedonia in schizophrenia: a review of assessment strategies. *Schizophrenia bulletin*, 32(2):259–273, 2005.

Hosking, J. G., Cocker, P. J., und Winstanley, C. A. Dissociable contributions of anterior cingulate cortex and basolateral amygdala on a rodent cost/benefit decision-making task of cognitive effort. *Neuropsychopharmacology*, 39(7):1558, 2014a.

Hosking, J. G., Cocker, P. J., und Winstanley, C. A. Dissociable contributions of anterior cingulate cortex and basolateral amygdala on a rodent cost/benefit decision-making task of cognitive effort. *Neuropsychopharmacology*, 39(7):1558, 2014b.

Huber, O. Das psychologische Experiment: Eine Einführung (5. überarb. Aufl.). *Bern: Hans Huber/Hogrefe*, 2009.

Hughes, D. M., Yates, M. J., Morton, E. E., und Smillie, L. D. Asymmetric frontal cortical activity predicts effort expenditure for reward. *Social Cognitive and Affective Neuroscience*, 10(7):1015–1019, 2014.

Hunton, J. E. und McEwen, R. A. An assessment of the relation between analysts' earnings forecast accuracy, motivational incentives and cognitive information search strategy. *Accounting Review*, 10(7):497–515, 1997.

Hussy, W., Schreier, M., und Echterhoff, G. *Forschungsmethoden in Psychologie und Sozialwissenschaften für Bachelor*. 2. Aufl., Berlin, Springer-Verlag, 2012.

Jeschke, B. G. *Entscheidungsorientiertes Management: Einführung in eine konzeptionell fundierte, pragmatische Entscheidungsfindung*. Berlin, Walter de Gruyter GmbH & Co KG, 2017.

Jia, Y., LENT, L. V., und Zeng, Y. Masculinity, testosterone, and financial misreporting. *Journal of Accounting Research*, 52(5):1195–1246, 2014.

Kable, J. W. The cognitive neuroscience toolkit for the neuroeconomist: A functional overview. *Journal of Neuroscience, Psychology, and Economics*, 4(2):63, 2011.

Kahneman, D. und Tversky, A. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291, 1979.

Kenning, P. *Consumer Neuroscience: Ein transdisziplinäres Lehrbuch*. Kohlhammer Verlag, 2014.

Kida, T. E., Moreno, K. K., und Smith, J. F. The influence of affect on managers' capital-budgeting decisions. *Contemporary Accounting Research*, 18 (3):477–494, 2001.

Kim, J., Hatcher, C., und Newton, C. Professional identity of management accountants: leadership in changing environments. *Management Accounting Research*, 18(3):477–494, 2012.

Kimberley, T. J. und Lewis, S. M. Understanding neuroimaging. *Physical therapy*, 87(6):670–683, 2007.

Kligman, A. M. Personal perspectives on bioengineering and the skin: The successful past and the brilliant future. *Handbook of non-invasive methods and the skin, Boca Raton*, pages 3–8, 2006.

Knutson, B., Rick, S., Wimmer, G. E., Prelec, D., und Loewenstein, G. Neural predictors of purchases. *Neuron*, 53(1):147–156, 2007.

Lashley, K. S. Basic neural mechanisms in behavior. *Psychological Review*, 37 (1):1, 1930.

Laux, H., Gillenkirch, R. M., und Schenk-Mathes, H. Y. Entscheidungstheorie (9., vollst. überarb. Aufl., Berlin), 2014.

Lechman, E. und Christiansen, B. *Neuroeconomics and the Decision-making Process*. IGI Global, Hersehy, 2016.

Lee, N., Broderick, A. J., und Chamberlain, L. What is 'neuromarketing'? a discussion and agenda for future research. *International Journal of Psychophysiology*, 63(2):199–204, 2007.

Lieberman, M. D. Social cognitive neuroscience: a review of core processes. *Annual Review of Psychology*, 58:259–289, 2007.

Locke, H. S. und Braver, T. S. Motivational influences on cognitive control: behavior, brain activation, and individual differences. *Cognitive, Affective, & Behavioral Neuroscience*, 8(1):99–112, 2008.

Loewenstein, G., Rick, S., und Cohen, J. D. Neuroeconomics. *Annual Review of Psychology*, 59:647–672, 2008.

Logothetis, N. K. und Pfeuffer, J. On the nature of the bold fmri contrast mechanism. *Magnetic Resonance Imaging*, 22(10):1517–1531, 2004.

Lorensen, T. D. und Dickson, P. Quantitative eeg normative databases: A comparative investigation. *Journal of Neurotherapy*, 7(3-4):53–68, 2003.

Lubow, R., Rifkin, B., und Alek, M. The context effect: The relationship between stimulus preexposure and environmental preexposure determines subsequent learning. *Journal of Experimental Psychology: Animal Behavior Processes*, 2(1):38–47, 1976.

Luck, S. J. *An introduction to the event-related potential technique*. 2. Aufl., Cambridge, Mit Press, 2014.

Maas, V. S., Van Rinsum, M., und Towry, K. L. In search of informed discretion: An experimental investigation of fairness and trust reciprocity. *The Accounting Review*, 87(2):617–644, 2011.

Mehta, P. H., Jones, A. C., und Josephs, R. A. The social endocrinology of dominance: basal testosterone predicts cortisol changes and behavior following victory and defeat. *Journal of Personality and Social Psychology*, 94(6):1078–1093, 2008.

Meyer, B. D. Natural and quasi-experiments in economics. *Journal of business & economic statistics*, 13(2):151–161, 1995.

Michel, C. M. und Murray, M. M. Towards the utilization of eeg as a brain imaging tool. *Neuroimage*, 61(2):371–385, 2012.

Moosbrugger, H. und Kelava, A. *Testtheorie und Fragebogenkonstruktion*. 2. Aufl., Berlin u.a., Springer, 2012.

Mukamel, R. und Fried, I. Human intracranial recordings and cognitive neuroscience. *Annual review of psychology*, 63:511–537, 2012.

Nilsson, A., Bergquist, M., und Schultz, W. P. Spillover effects in environmental behaviors, across time and context: a review and research agenda. *Environmental Education Research*, 23(4):573–589, 2017.

Nooraie, M. Factors influencing strategic decision-making processes. *Inter-*

national Journal of Academic Research in Business and Social Sciences, 2 (7):405–429, 2012.

Oblak, K., Ličen, M., und Slapničar, S. The role of cognitive frames in combined decisions about risk and effort. *Management Accounting Research*, 39:35–46, 2018.

Ostrom, T. M., Betz, A. L., und Skowronski, J. J. Cognitive representation of bipolar survey items. In *Context effects in social and psychological research*. New York, Springer: 297-311, 1992.

Peecher, M. E., Solomon, I., und Trotman, K. T. An accountability framework for financial statement auditors and related research questions. *Accounting, Organizations and Society*, 38(8):596–620, 2013.

Penton-Voak, I. S., Pound, N., Little, A. C., und Perrett, D. I. Personality judgments from natural and composite facial images: More evidence for a "kernel of truth" in social perception. *Social Cognition*, 24(5):607–640, 2006.

Pieters, R. und Wedel, M. Goal control of attention to advertising: The yarbus implication. *Journal of Consumer Research*, 34(2):224–233, 2007.

Pochon, J., Levy, R., Fossati, P., Lehericy, S., Poline, J., Pillon, B., Le Bihan, D., und Dubois, B. The neural system that bridges reward and cognition in humans: an fmri study. *Proceedings of the National Academy of Sciences*, 99(8):5669–5674, 2002.

Poldrack, R. A. Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, 10(2):59–63, 2006.

Pound, N., Penton-Voak, I. S., und Surridge, A. K. Testosterone responses to competition in men are related to facial masculinity. *Proceedings of the Royal Society B: Biological Sciences*, 276(1654):153–159, 2008.

Prelec, D. und Loewenstein, G. The red and the black: Mental accounting of savings and debt. *Marketing Science*, 17(1):4–28, 1998.

Raghubir, P. und Srivastava, J. Monopoly money: The effect of payment coupling and form on spending behavior. *Journal of experimental psychology: Applied*, 14(3):213–225, 2008.

Raichle, M. E. Behind the scenes of functional brain imaging: a historical and physiological perspective. *Proceedings of the National Academy of Sciences*, 95(3):765–772, 1998.

Rayner, K. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3):372–422, 1998.

Reichmann, T. *Controlling mit Kennzahlen und Managementberichten*, volume 5. Vahlen-Verlag, München, 1997.

Reimann, M. und Weber, B. *Neuroökonomie: Grundlagen - Methoden - Anwendungen*. Gabler Verlag, Wiesbaden, 2011.

Renner, K.-H., Heydasch, T., und Ströhlein, G. *Forschungsmethoden der Psychologie*. 1. Aufl., Wiesbaden, Springer, 2012.

Riedl, R. und Léger, P.-M. Tools in neurois research: an overview. In *Fundamentals of NeuroIS*. Berlin, Springer: 47-72, 2016.

Roth, G. *Fühlen, Denken, Handeln*. Frankfurt am Main, 2003.

Salamone, J. D., Cousins, M. S., und Bucher, S. Anhedonia or anergia? effects of haloperidol and nucleus accumbens dopamine depletion on instrumental response selection in a t-maze cost/benefit procedure. *Behavioural Brain Research*, 65(2):221–229, 1994.

Salmelin, R. und Baillet, S. Electromagnetic brain imaging. *Human Brain Mapping*, 30(6):1753–1757, 2009.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., und Cohen, J. D. The neural basis of economic decision-making in the ultimatum game. *Science*, 300(5626):1755–1758, 2003.

Sarris, V. und Reiß, S. *Kurzer Leitfaden der Experimentalpsychologie*. München, Pearson Studium, 2005.

Schilke, O. und Reimann, M. Neuroökonomie: Grundverständnis, Methoden und betriebswirtschaftliche Anwendungsfelder. *Journal für Betriebswirtschaft*, 57(3-4):247–262, 2007.

Schmid, H. Qualitative research and occupational therapy. *American Journal of Occupational Therapy*, 35(2):105–106, 1981.

Schröger, E. *Biologische Psychologie*. Wiesbaden, Springer, 2011.

Schwarz, E. *Neuro-Advertising: Gehirngerechte Werbung für mehr Erfolg in Ihrem Markt*. Springer, Wiesbaden, 2018.

Schwenk, C. R. Cognitive simplification processes in strategic decision-making. *Strategic Management Journal*, 5(2):111–128, 1984.

Shanks, N., Greek, R., und Greek, J. Are animal models predictive for humans? *Philosophy, Ethics, and Humanities in Medicine*, 4(1):1–20, 2009.

Shimojo, S., Simion, C., Shimojo, E., und Scheier, C. Gaze bias both reflects and influences preference. *Nature Neuroscience*, 6(12):1317–1322, 2003.

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., und Frith, C. D. Empathic neural responses are modulated by the perceived fairness of others. *Nature*, 439(7075):466–469, 2006.

Soman, D. Effects of payment mechanism on spending behavior: The role of rehearsal and immediacy of payments. *Journal of Consumer Research*, 27(4):460–474, 2001.

Starck, J. M. *Peer Review für wissenschaftliche Fachjournale*. Wiesbaden, Springer, 2018.

Steiner, A. P. und Redish, A. D. Behavioral and neurophysiological correlates of regret in rat decision-making on a neuroeconomic task. *Nature Neuroscience*, 17(7):995–1002, 2014.

Sternberg, R. J. Testing and cognitive psychology. *American Psychologist*, 36 (10):1181–1189, 1981.

Stewart, N., Hermens, F., und Matthews, W. J. Eye movements in risky choice. *Journal of Behavioral Decision Making*, 29(2-3):116–136, 2016.

Stroebe, W. *Grundlagen der Sozialpsychologie 1*. Stuttgart, Klett-Cotta, 1980.

Sun, Q., Gu, S., und Yang, J. Context and time matter: Effects of emotion and motivation on episodic memory overtime. *Neural Plasticity*, 2018(1): 1–13, 2018.

Swait, J., Adamowicz, W., Hanemann, M., Diederich, A., Krosnick, J., Layton, D., Provencher, W., Schkade, D., und Tourangeau, R. Context dependence and aggregation in disaggregate choice analysis. *Marketing Letters*, 13(3): 195–205, 2002.

Sweeny, R. B. MAP Committee Promulgates Definition of Management Accounting. *Thomas, William E. (Hrsg.), Readings in Cost Accounting Budgeting and Control, 1983, 6. Aufl., Cincinnati*, pages 466–469, 1983.

Taylor, S. F., Welsh, R. C., Wager, T. D., Phan, K. L., Fitzgerald, K. D., und Gehring, W. J. A functional neuroimaging study of motivation and executive function. *Neuroimage*, 21(3):1045–1054, 2004.

Telpaz, A., Webb, R., und Levy, D. J. Using eeg to predict consumers' future choices. *Journal of Marketing Research*, 52(4):511–529, 2015.

Toga, A. W., Frackowiak, R., und Mazziotta, J. C. Neuroimage anniversary issue. *NeuroImage*, 61(2):323, 2012.

Töpfer, A. *Erfolgreich Forschen: Ein Leitfaden für Bachelor-, Master-Studierende und Doktoranden*. Wiesbaden, Springer-Verlag, 2012.

Treadway, M. T., Peterman, J. S., Zald, D. H., und Park, S. Impaired effort allocation in patients with schizophrenia. *Schizophrenia Research*, 161 (2-3):382–385, 2015.

Tuthill, J. The rat makes the decisions around here. *Journal of Experimental Biology*, 218(9):1280–1280, 2015.

Tversky, A. und Kahneman, D. Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131, 1974.

Tversky, A. und Kahneman, D. The framing of decisions and the psychology of choice. *Science*, 211(4481):453–458, 1981.

Verplaetse, J., Vanneste, S., und Braeckman, J. You can judge a book by its cover: the sequel.: A kernel of truth in predictive cheating detection. *Evolution and Human Behavior*, 28(4):260–271, 2007.

VHB JOURQUAL 3. [19.11.2018], 2018. URL `http://vhbonline.org/vhb4you/jourqual/vhb-jourqual-3/teilrating-rech/`.

Volk, S. und Köhler, T. Brains and games: Applying neuroeconomics to organizational research. *Organizational Research Methods*, 15(4):522–552, 2012.

Vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., Cleven, A., et al. Reconstructing the giant: on the importance of rigour in documenting the literature search process. In *Proceedings of the 17th European Conference on Information Systems, Vol. 161: 1-14*, 2009.

Wall, F. und Schröder, R. W. Zwischen shareholder value und stakeholder value: Neue herausforderungen für das controlling?! *Controlling zwischen Shareholder Value und Stakeholder Value: Neue Anforderungen, Konzepte und Instrumente, München*, pages 3–18, 2009.

Walton, M., Kennerley, S., Bannerman, D., Phillips, P., und Rushworth, M. F. Weighing up the benefits of work: behavioral and neural analyses of effort-related decision making. *Neural Networks*, 19(8):1302–1314, 2006.

Wang, B. Analysis of Corporate Social Responsibility Decision-making Behavior Based on Cognitive Neuroscience. *Neuro Quantology*, 18(6):326–333, 2018.

Wardle, M. C., Treadway, M. T., und de Wit, H. Caffeine increases psychomotor performance on the effort expenditure for rewards task. *Pharmacology Biochemistry and Behavior*, 102(4):526–531, 2012.

Waymire, G. B. Exchange guidance is the fundamental demand for accounting. *The Accounting Review*, 84(1):53–62, 2009.

Waymire, G. B. Neuroscience and ultimate causation in accounting research. *The Accounting Review*, 89(6):2011–2019, 2014.

Webster, J. und Watson, R. T. Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, 26(2):xiii–xxiii, 2002.

Westlund, K. Training laboratory primates–benefits and techniques. *Primate Biology*, 2(1):119–132, 2015.

Williamson, O. E. Opportunism and its critics. *Managerial and Decision Economics*, 14(2):97–107, 1993.

Windeler, J. Externe validität. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 102(4):253–259, 2008.

Wright, N. D., Bahrami, B., Johnson, E., Di Malta, G., Rees, G., Frith, C. D., und Dolan, R. J. Testosterone disrupts human collaboration by increasing egocentric choices. *Proceedings of the Royal Society B: Biological Sciences*, 279(1736):2275–2280, 2012.

Zeelenberg, M. Anticipated regret, expected feedback and behavioral decision making. *Journal of Behavioral Decision Making*, 12(2):93–106, 1999.

Zellermayer, O. The pain of paying. *unpublished dissertation, Department of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, PA*, 1996.

Zingales, L. The future of securities regulation. *Journal of Accounting Research*, 47(2):391–425, 2009.

Zuckerman, M. und Kuhlman, D. M. Personality and risk-taking: common bisocial factors. *Journal of Personality*, 68(6):999–1029, 2000.

Junior Management Science 4(3) (2019) 422-432

Junior Management Science

journal homepage: www.jums.academy

Measuring the Impact of Carbon Emissions on Firm Value Using Quantile Regression

Lukas Ferner

Universität Augsburg

Abstract

A fundamental transformation of the global economy towards a low-carbon economy is inevitable in order to achieve the climate targets set by the United Nations. Hence, it becomes increasingly important to understand how firm level carbon mitigation affects the value of a company. The purpose of this thesis is not only to estimate the average relationship between carbon emissions and firm value but to investigate whether this relationship is heterogeneous and thus whether the effect of carbon emission on firm value depends on the value of the respective company. A quantile regression approach with firm value measured as Tobin's Q as the dependent variable is applied. The estimation outcomes clearly indicate that higher carbon emissions reduce firm value for all quantiles. However, the extent of the effect depends strongly on the value of the respective company suggesting that the value-enhancing effect of reduced carbon emissions is higher for firms with relatively high firm value.

Keywords: carbon emission; firm value; quantile regression.

1. Introduction

In 2017, the annual average amount of carbon dioxide in earth's atmosphere reached an all-time peak of more than 400 parts per million (ppm). This corresponds to an alarming increase in atmospheric carbon dioxide concentration of about 25% in the last 50 years (Lindsey, 2018). Evidence suggests that this man-made increase has contributed significantly to progressive global warming (Williams et al., 2017) and thereby intensified the associated problems in financial, social, and environmental aspects on a global scale (United States Environmental Protection Agency, 2017). To avoid a worsening of the impacts and to reduce the risks of global warming, 195 nations devoted themselves to hold the increase in the global average temperature well below 2°C compared to pre-industrial levels by reducing greenhouse gas emissions, especially carbon dioxide emissions (United Nations, 2015). However, in a very recent report, the intergovernmental panel on climate change (IPCC) shows that the measures taken so far are not sufficient. The report finds that limiting global warming to 1.5°C - 2°C would still require a reduction of carbon emissions of about 45% by 2030, reaching net zero by 2050. The achievement of this ambitious objective calls for far-reaching and rapid changes in every aspect of human activities (IPCC, 2018). Consequently, firm level adaption and carbon emission mitigation are essential, because the corporate sector is responsible for a significant amount of overall carbon emissions. Hence, it is most likely that companies will face increased shareholder and public pressure to reduce carbon emissions as well as more stringent climate change regulations (Aggarwal and Dow, 2011). Moreover, the direct financial impact of carbon emissions becomes increasingly relevant due to the resolutions made during the world climate conference in Paris 2015 (Union Investment, 2016). These developments mean that firm level reduction of carbon emissions becomes increasingly important, not only for companies but for all parts of an economy. In order to take appropriate measures, corporate managers, investors and regulators have to deal with how the inevitable transition to a low-carbon economy and the associated mitigation of carbon emissions affects a firm's value and financial performance (Lee and Min, 2015; Union Investment, 2016). Therefore, it is essential to provide these decision-makers with an accurate and reliable estimation of the relationship. In prior research, it is generally suggested that environmental and economic performance are positively correlated. Nevertheless, as only mean regression techniques have been applied to investigate this issue, the relation between carbon emissions and firm value is solely estimated with respect to the 'average firm' (Mosteller and Tukey, 1977). This lead to the presentation of an incomplete picture of the relationship (Koenker and Bassett Jr, 1978). By applying the concept of quantile regression, this study explores the possibility of dif-

DOI: https://doi.org/10.5282/jums/v4i3pp422-432

ferent effects of carbon emissions on firm value depending on the value of the respective firm. Hence, more detailed results and more reliable estimators (Koenker and Hallock, 2001) on the relationship are provided.

The remainder of this study is structured as follows: In section two, a short overview of the current state of research on the relation between carbon emissions and firm value is given and the hypothesis to be tested in this study is developed. In section three, the theoretical concept of quantile regression is presented. In section four, the examined data set is introduced. In section five, a mean and quantile regression analysis is conducted and the results are presented. In section six, the results are discussed and possible implications are presented. Moreover, limitations of the study and suggestions for further research are provided. In the last section, the study is briefly summarized.

2. Literature review and development of hypothesis

2.1. Overview of the current state of research

As was to be expected, a considerable amount of work has been done to analyse the impact of a firm's environmental performance on its valuation. Most studies examined the relationship based on the hypothesis, that better (worse) environmental performance increases (decreases) firm value or financial performance. Indicators most often used as a proxy for environmental performance include the amount of carbon emissions, greenhouse gas emissions and aggregated environmental performance indices (e.g. Aggarwal and Dow, 2011; Al-Najjar and Anfimiadou, 2012; Nishitani and Kokubu, 2012; Matsumura et al., 2014). As carbon dioxide is reported to be the principal source of global warming, it is directly associated with the most impending global environmental issue (Lee and Min, 2015). Therefore, carbon emission is used as the indicator for a firm's environmental performance in this study. Regardless of the specific definition of environmental performance, an overwhelming majority of research on this issue found a positive impact of environmental activities on firm value. Especially with increased public perception of the negative consequences of global warming (Capstick et al., 2015), a discernible trend in recent research, suggesting that companies can improve its financial performance by reducing emissions, can be observed. For example, Matsumura et al. (2014) conducted a study with a sample of S&P 500 firms and found that, on average, firm value decreases by $212.000 for every additional thousand metric tons of carbon dioxide emitted. In another study on South African companies, Ganda and Milondzo (2018) focused on the impact of carbon emissions not on firm value but on corporate financial performance in general using multiple regression techniques. They found strong evidence that supports the idea of a negative relation between carbon emissions and financial performance. In a similar study on Japanese manufacturing firms, Nishitani and Kokubu (2012) exploited a random effects model to investigate the relationship between a firm's greenhouse gas emissions and firm value measured as Tobin's Q. Their observations were in line with former results as they found that firms are more likely to increase in value when reducing greenhouse gas emissions. King and Lenox (2008) applied a panel data regression to analyse the relation between carbon emissions and firm value with a sample of 614 US manufacturing firms. They found that lower emissions are significantly associated with higher Tobin's Q results. Furthermore, several other studies including Aggarwal and Dow (2011) and Al-Najjar and Anfimiadou (2012) found evidence of a negative relationship between carbon emissions and firm value in particular or a positive relationship between environmental and economic performance in general.

A common approach to explain why environmental activities improve firm value is based on the assumption that stakeholders and investors consider environmental performance as a form of intangible value. This is because successful emission mitigation decreases the exposure to global warming risks and generates new profit opportunities. Thus, a competitive advantage over rivals in a low-carbon future is generated (Nishitani and Kokubu, 2012). Moreover, especially with increased public perception of global warming issues, the positive reputational effect of green investment activities becomes more important, and thus more valuable. Besides, increased energy efficiency (Sprinkle and Maines, 2010) and the possibility to charge higher prices for environmentally friendly products (Aggarwal and Dow, 2011) are presumed to be firm value-enhancing effects of increased environmental performance.

2.2. Purpose of this bachelor's thesis

Although the above-mentioned studies generally suggest that environmental activities and performance are positively correlated with firm value, the informativeness of this major finding is limited by the fact that the relation is always estimated by applying different mean regression techniques (e.g. Aggarwal and Dow, 2011; Al-Najjar and Anfimiadou, 2012; Matsumura et al., 2014; Ganda and Milondzo, 2018). Consequently, prior research focused on measuring the average effect, which is, according to Koenker and Hallock (2001), likely to provide an incomplete picture of a relation between two variables. It is not considered, that this average effect of carbon emissions on firm value may not necessarily be representative for all firms. This reduces the reliability of the provided results (Koenker and Hallock, 2001).

Due to the lack of research on this issue, the focus of this study is to shed light on the heterogeneity of the effect of carbon emissions on firm value. Thereby, a quantile regression model as introduced by Koenker and Bassett Jr (1978) is applied and the following hypothesis is to be tested:

> H: The effect of carbon emissions on firm value depends significantly on the value of the respective firm, and thus the average effect is not informative.

3. Theoretical background on quantile regression

3.1. Motivation of quantile regression

To provide an estimation on the relationship between a set of independent variables and a dependent variable, standard ordinary least squares regression is one of the most popular statistical methods (Huang et al., 2017). However, this approach, as well as other mean regression techniques, gives a rather incomplete picture of the covariate effects just as a sample's mean gives an incomplete picture of a single distribution (Mosteller and Tukey, 1977). This is because mean regression models quantify the effects of the explanatory variables only at the conditional mean and assume that this effect is constant, and thus representative, for the entire conditional distribution of the dependent variable (Huang et al., 2017). If one expects the effects of the explanatory variables on the dependent variable to be heterogenous, this average effect is not adequate (Koenker and Hallock, 2001; Huang et al., 2017). As stated by Mosteller and Tukey (1977), it gives an incomplete picture and, moreover, might lead to wrong conclusions and less informative results (Schulze, 2004). Quantile regression addresses this issue as it allows to evaluate the covariate effects separately for each quantile along the entire conditional distribution of the dependent variable instead of only the effects at the conditional mean (Huang et al., 2017). Hence, quantile regression gives a far broader picture of the covariate effects (Koenker and Hallock, 2001) and provides the possibility to detect heteroscedasticity in the examined data (Schulze, 2004). However, if the relation between an dependent and an independent variable is highly homogenous amongst the entire conditional distribution, quantile regression does not provide any added value (Koenker and Hallock, 2001). Beside the possibility to capture heterogenous effects, quantile regression has other favourable characteristics. It is well known that, when asymmetries and heavy tails exist, the sample median provides a better summary of centrality than the mean (Koenker and Hallock, 2001). Consequently, compared to standard mean models, quantile regression is far more robust to outliers (Huang et al., 2017). Moreover, there is no need for an assumption about the parametric distribution of the dependent variable and the constant variance of the observation-specific error terms (Huang et al., 2017).

3.2. Cross-sectional quantile regression

The original quantile regression model as introduced by Koenker and Bassett Jr (1978) is based on

$$y_i = x_i\beta_\tau + \epsilon_i \; ; \; i = 1, \ldots, n \; ; \; \tau \in (0;1) \tag{1}$$

where x denotes a vector of explanatory variables, y is the dependent variable and ϵ presents the observation-specific disturbance term. β_τ denotes the vector of quantile-specific regression parameters to be estimated for the τ-th quantile of the conditional distribution of the dependent variable (Koenker and Hallock, 2001). The estimation of the parameters is based on solving a least asymmetric absolute deviation problem:

$$\begin{aligned}\hat{\beta}_\tau = & argmin_{\beta_\tau}\Big(\sum_{i\in\{i|y_i\geq x_i\beta_\tau\}} \tau * |y_i - x_i\beta_\tau| + \\ & \sum_{i\in\{i|y_i<x_i\beta_\tau\}} (1-\tau) * |y_i - x_i\beta_\tau|\Big)\end{aligned} \tag{2}$$

To paraphrase it, all observations above the hyperplane $x_i\beta_\tau$ are multiplied with τ, all observations below with $(1-\tau)$. Then, β_τ shall be chosen such that the sum of these asymmetrically weighted absolute deviations is minimized. This minimization problem can mostly be solved efficiently by linear programming methodologies (Schulze, 2004). As the problem depends on τ, the solution to the problem yields a distinct set of regression coefficients for each desired quantile level. Hence, instead of estimating solely one conditional mean function as in standard mean regression models, a quantile regression model provides a set of conditional quantile functions (Schulze, 2004). The τ-th conditional quantile of x_i with respect to x_i can then be written as

$$Q_\tau(y|x) = x\hat{\beta}_\tau \tag{3}$$

Thereby, the obtained τ-specific regression coefficients $\hat{\beta}_\tau$ can be interpreted as the effect of the explanatory variables at the τ-th quantile of the conditional distribution of the dependent variable (Koenker and Bassett Jr, 1978).

3.3. Quantile regression on panel data

As the data set that is exploited in the following empirical analysis has both a cross-sectional and a time dimension, the cross-sectional quantile regression approach of Koenker and Bassett Jr (1978) ought to be extended in order to capture unobserved individual-specific effects (Koenker, 2004). However, models combining the quantile regression framework with panel data analysis are based on very strong assumptions that are rarely satisfied. Especially with a rather small amount of time series, as in the present data, it is difficult to provide any meaningful results with most approaches (Powell, 2015). Besides, the interpretation of the estimation outcomes of most approaches varies significantly compared to cross-sectional quantile regression coefficients (Powell, 2016). Due to these difficulties in estimating a quantile regression model for panel data, the time dimension of the data set used in this study is eliminated in the following empirical analysis. Thus, the cross-sectional quantile regression approach introduced earlier in this section can be applied. Nevertheless, the combination of quantile regression and panel data analysis provides some favourable features as it allows to control for unobserved individual-specific effects and to expose heterogenous covariate effects simultaneously (Huang et al., 2017). Therefore, a very recent quantile regression approach for panel data introduced by Powell (2015) is attached to the study.

4. Variables & sample research

4.1. Data sources

The following empirical models are based on firm observations for the years 2010 to 2016. Financial data were obtained from the Thomson Reuters Datastream and are measured in U.S. dollars. Carbon emission data were collected from two different databases: the Carbon Disclosure Project (CDP) and Thomson Reuters ESG. The amount of carbon emissions for each firm and year is expressed in tonnes. The disclosure of data on carbon emissions to these databases is voluntary. For firms that did not disclose data voluntarily, the amount of carbon emissions was estimated if reliably possible (CDP Worldwide, 2017). Due to this estimation approach, some of the obtained measurements of a firm's carbon emission do differ slightly depending on the respective database. Thus, when data on carbon emission for one firm and year are available in more than one database, the average measurement is used.

4.2. Independent variables

The relevant variable of interest in this study is carbon emission. For each firm and year, overall carbon emissions, Scope 1 carbon emissions and Scope 2 carbon emissions are reported. Whereas Scope 1 refers to carbon emissions from sources that are controlled or owned by the company (direct carbon emissions), Scope 2 accounts for carbon dioxide that is emitted by generation of purchased electricity consumed by the firm (indirect carbon emissions). Overall carbon emission accounts for both Scope 1 and Scope 2, as well as for all other carbon emissions not covered by these two categories (World Resources Institute, 2014). To provide a picture as detailed as possible, the effect of Scope 1, Scope 2 and overall carbon emissions on firm value is estimated separately. Hence, at each step of the following empirical analysis, three models are estimated. The first model estimates the effect of direct carbon emissions on firm value. The second model focuses on the relation between indirect carbon emissions and firm value. In the third model, the relation between a firm's overall carbon emissions and firm value is analysed. For the purpose of improved representation of the estimation outcomes, the amount of a firm's carbon emissions in each category is divided by 1,000,000.

4.3. Dependent variable

To follow prior research in this area, firm value measured as Tobin's Q denotes the dependent variable. A firm's Tobin's Q is calculated as the ratio of the market value of assets to the replacement value of assets. Market value of assets is calculated as the sum of market value of equity and market value of liabilities. The replacement value of assets is calculated as the sum of book value of equity and book value of liabilities (Vermunt, 2013). Hence, the formulation to calculate a firm's Tobin's Q is:

$$Q_i = (\text{market value of equity} + \text{market value of liabilities}) / (\text{book value of equity} + \text{book value of liabilities}) \quad (4)$$

In literature, some more elaborate calculation approaches may be found. However, the simplified approach used in this study significantly reduces the computational effort. It is widely used in recent research and presents a sufficient approximation of most more complex approaches (Dowell et al., 2000; King and Lenox, 2008). The use of Q as the measurement of firm value allows to directly compare accounting data and financial valuation data. This offers the opportunity to expose the market evaluation of management performance. Moreover, it allows considering both a firm's results produced in the past and potential future growth opportunities (Lindenberg and Ross, 1981). A value of Q above 1 suggests that an investment in the respective firm would be profitable, because the value of the capital investments exceeds the costs (King and Lenox, 2008). It indicates good management performance as the outputs (market value) exceed the inputs (book value) (Lindenberg and Ross, 1981). A Q value below 1 indicates that the market assumes the future cash flows provided by the firm to be lower than the amount of money invested in its assets (King and Lenox, 2008).

4.4. Control variables

Several control variables are considered in this study which are commonly included in financial performance analysis and known to impact Tobin's Q (e.g. Konar and Cohen, 2001; King and Lenox, 2008; Aggarwal and Dow, 2011; Lee and Min, 2015; Iskandar, 2016). The control variables are 1) capital intensity calculated as the ratio of capital expenditures to net sales, 2) growth rate calculated as the percentage change in net sales, 3) leverage measured as the ratio of total debt to total assets, 4) profitability calculated as the ratio of net income to total assets, 5) R&D intensity calculated as the ratio of R&D expenditures to total assets and 6) liquidity measured as the ratio of cash to total assets. Table 1 provides an overview of all variables included in the following regression analysis.

4.5. Sample

As a cross-sectional regression analysis is conducted in the following empirical section, the time dimension of the data was eliminated. Thereby, for each firm and variable, the average value across all years was calculated. For a firm to be included in the sample, Tobin's Q, measurements of all three carbon emission categories as well as the introduced control variables must be available for at least two years in the period of 2010-2016. To ensure that the regression results are not unduly sensitive to outliers, all variables are 'winsorized'. Thereby, 2.5% of the measurements are modified at each tail of the respective distribution (Cox, 1998). After excluding firms that lacked sufficient financial or carbon emission data, a sample of 3,570 firms remained. Testing for multicollinearity of the explanatory variables by calculating the correlation

Table 1: Overview of all variables and the respective definition

Dependent Variable	
Tobin's Q	Market value over replacement value of assets
Independent Variables	
Scope 1	Direct carbon emissions divided by 1 million
Scope 2	Indirect carbon emissions divided by 1 million
Overall	Overall carbon emissions divided by 1 million
Control Variables	
Capital Intensity	Capital expenditures over net sales
Growth Rate	Percentage change in net sales
Leverage	Total debts over total assets
Profitability	Net Income over total assets
R&D Intensity	R&D expenditures over total assets
Liquidity	Cash over total assets

matrix and the variance inflation factors, which are presented in the appendix, did not reveal any problem. In Table 2, the descriptive statistics of all variables employed for the empirical models are displayed.

5. Empirical models

5.1. Classical mean regression analysis

Before conducting a quantile regression analysis to address potential heterogenous effects, standard ordinary least squares regression is applied. Hence, the average relationship between the three categories of carbon emission (Scope 1, Scope 2, Overall) and firm value is estimated (Mosteller and Tukey, 1977). This allows to give a reference point for the following quantile analysis and to obtain an initial overview of the relationship. Moreover, it is evaluated whether the predominant finding of a negative average effect of carbon emissions on firm value in recent research can be supported. Thereby, one regression model is estimated for each of the three carbon emission categories. Each model includes Tobin's Q as the dependent variable and the respective carbon emission measurement and all control variables as the explanatory variables. For all three models, an adjusted R2 value of approximately 0.4 was observed. The F-test approved the overall model significance for all mean regression models. The models were re-estimated using robust standard errors. However, the outcomes were not significantly different from the results presented below. In the first model, a firm's direct carbon emissions are included as the relevant variable of interest. Table 3 presents the estimation results.

The major finding of the first model is that the mean effect of direct carbon emissions on firm value is found to be negative and statistically significant. This indicates that firms are, on average, more likely to increase Tobin's Q when reducing direct carbon emissions. In contrast to Scope 1 carbon emissions, all control variables impact Tobin's Q positively. According to Bartram et al. (2011), profitability and growth are likely to increase firm value, because they lead to higher cash flows to equity holder. Du et al. (2016) found that sufficient liquidity also increases firm value measured as Tobin's Q, because it increases the financial scope and reduces the risk of insolvency. High research and development intensity might promise future comparative advantages and is therefore positively linked with firm value (Gupta et al., 2017; Du et al., 2016). Leverage and capital intensity were not found to impact firm value significantly.

The second mean regression model estimates the average relation between indirect carbon emissions and firm value. The observed regression coefficients as well as the associated p-values are presented in Table 4.

The estimation outcomes of the second model show that the coefficient carried by Scope 2 is negative and statistically significant. Thus, the model suggests that companies can increase firm value by reducing indirect carbon emissions. As expected, the estimated coefficients of the control variables do not change remarkable in sign, value or significance when compared to the first model.

Table 5 provides information on how firm value is affected by a company's overall carbon emissions. These include Scope 1 and Scope 2 carbon emissions as well as all other carbon emissions not covered by these two categories.

As was to be expected with respect to the first two models, the coefficient of the relevant explanatory variable shows that the effect of overall carbon emissions on firm value is negative and statistically significant. Thus, one can conclude that a company is, on average, likely to increase firm value when reducing carbon emissions regardless of the specific source of emission. The effects of the control variables on firm value are similar to those observed in the first two models.

Table 2: Summary statistics for sample companies

Variable	Mean	Median	Standard Dev.
Tobin's Q	2.110	1.605	1.363
Scope 1	0.792	0.035	2.256
Scope 2	0.215	0.048	0.384
Overall	1.105	0.105	2.779
Capital Intensity	0.115	0.044	0.241
Growth Rate	0.113	0.047	0.244
Leverage	0.217	0.199	0.171
Profitability	0.028	0.040	0.095
R&D Intensity	0.042	0.012	0.070
Liquidity	0.126	0.090	0.117

Table 3: Mean regression estimators – Direct Carbon Emission

Variable	Coefficient	p-value
Scope 1	- 0.0318	0.000
Capital Intensity	0.1194	0.133
Growth Rate	1.4040	0.000
Leverage	0.0128	0.911
Profitability	4.2904	0.000
R&D Intensity	8.4910	0.000
Liquidity	2.7320	0.000

Table 4: Mean regression estimators – Indirect Carbon Emission

Variable	Coefficient	p-value
Scope 2	- 0.3332	0.000
Capital Intensity	0.9426	0.233
Growth Rate	1.3722	0.000
Leverage	0.0541	0.637
Profitability	4.3020	0.000
R&D Intensity	8.4626	0.000
Liquidity	2.6616	0.000

Table 5: Mean regression estimators – Overall Carbon Emission

Variable	Coefficient	p-value
Overall	- 0.0282	0.000
Capital Intensity	0.1171	0.141
Growth Rate	1.3997	0.000
Leverage	0.0188	0.870
Profitability	4.2870	0.000
R&D Intensity	8.4820	0.000
Liquidity	2.7198	0.000

5.2. Quantile regression analysis

As described in section 3, the results obtained from the above conducted mean regression analysis provide only a rough summary of the effects of carbon emissions on firm value as only the effect at the condition mean is estimated (Koenker and Hallock, 2001). To expose potential heterogeneity of the effects, quantile regression estimators as introduced by Koenker and Bassett Jr (1978) are calculated. Analogue to the mean regression models, one quantile regression model is estimated for each carbon emission category. Thereby, the obtained quantile-specific regression coefficients indicate how the respective carbon emission category affects firm value at the τ-th quantile of the conditional distribution of firm value measurements (Koenker and Hallock, 2001). Tables 6-8 provide an extract of the estimation outcomes. For each emission category, the estimated re-

gression coefficient and the associated p-value are given for each decile of the conditional distribution. To stay within the scope of the paper, the coefficients carried by the control variables are not displayed at this point of the study. However, the complete estimation outcome including the coefficients of the control variables and the related p-values are attached to this study. To illustrate the statistical variation of the coefficients carried by the variables of interest along the entire conditional distribution of the dependent variable, a graphic display is provided. In each graph, the estimated quantile regression coefficients of the respective carbon emission category are plotted against the quantiles of the conditional distribution. The solid horizontal line denotes the mean regression estimate which does obviously not vary with the quantiles. The two dashed lines depict the conventional 95% confidence interval for the respective mean regression estimator.

Looking at the graphical depiction presented above reveals that carbon emissions from all sources affect firm value similarly along the conditional distribution. However, it should be noted that the coefficients obtained from the second model, referring to the impact of Scope 2 carbon emissions on firm value, are much higher at each specific decile than the estimators provided by the first and the third model. Nevertheless, at the bottom of the conditional distribution, the quantile estimators carried by the variable of interest are close to zero, but always negative, in all three models. The effect of direct carbon emissions on firm value at the lowest decile is not found to be statistically significant at the 10% level. From the conditional distribution's bottom until approximately the median, the quantile-specific coefficients in all models are roughly constant as they show only minor changes with a tendency to decrease (increase in absolute value). In contrast, at the right side of the conditional distribution, the coefficients carried by the independent variables diminish strongly reaching the lowest (highest absolute) value at the highest quantiles. Except the mentioned insignificance, all quantile estimators in all three models are significant at the 10% level.

However, the described shift of the quantile estimators along the conditional distribution of firm value measurements does not justify a quantile regression model as long as the results are not compared to the mean regression estimator, especially to its confidence interval (Koenker and Hallock, 2001). Hence, the main important result from this quantile regression analysis is that the obtained quantile-specific regression coefficients from all three models lie predominantly outside the 95% confidence interval of the mean regression estimator. This clearly indicates that the least squares confidence interval does a poor job representing the range of covariate effects along the conditional distribution (Koenker and Hallock, 2001). The average effect is neither representative for the quantiles below the median nor for the effects at top of the conditional distribution. Solely between approximately the sixth and the ninth decile, the quantile effects are adequately expressed by the mean regression models. This range is even smaller for the quantile effects provided by the second model. This clearly demonstrates that the quantile regression approach improves the informativeness and reliability of the results compared to the classical OLS regression. Moreover, it can be seen that the median estimator lies significantly above the mean estimator in all three models. This emphasizes the robustness of the quantile regression approach towards outliers. In contrast, the mean effect of carbon emissions on firm value is likely to be distorted by the very strong quantile effects at the top of the conditional distribution.

The coefficients carried by the control variables are, as well as in the mean regression analysis, mostly constant across the three models. However, the quantile regression approach revealed significant heterogeneity of the effects along the conditional distribution of firm value measurements for most of the control variables. Hence, it would be worth investigating the impact of financial performance indicators on firm value using quantile regression in further research.

6. Discussion

6.1. Comparison and interpretation of the regression results

The mean regression models support the trend in recent research suggesting that firms can, on average, increase firm value when reducing carbon emissions. The estimation outcomes clearly show that lower direct, indirect and overall carbon emissions are significantly associated with higher Tobin's Q results. Thus, the reduction of carbon emissions is likely to improve firm value regardless of the specific source of emission. Thereby, the value-enhancing effect of reduced Scope 2 carbon emissions is estimated to be stronger than the effect of reduced direct carbon emissions along the entire conditional distribution. Moreover, the effect of overall carbon emissions on firm value is lower than the effects of Scope 1 and Scope 2 carbon emissions. This may suggest that all carbon emissions not covered by Scope 1 and Scope 2, but included in the overall emission measurement, have a lower negative or even positive impact on firm value. As well as the mean regression models, the quantile approach generally suggests that lower carbon emissions are significantly associated with higher firm value measurements. Regardless of the specific source of emission and the specific quantile of the conditional distribution of the dependent variable, carbon emission reduction is likely to improve firm value. However, the extent to which a firm can benefit from carbon emission mitigation varies significantly along the conditional distribution. The models suggest that firms with relatively low Tobin's Q results cannot improve firm value by reducing carbon emissions as much as suggested by the average effect. In contrast, for firms at the top of the conditional distribution of firm value measurements, the value-enhancing effect of reduced carbon emissions is significantly higher and therefore strongly underestimated by the average effect. This applies to Scope 1, Scope 2 and overall carbon emissions.

As can be seen from the empirical results, the underlying hypothesis of this study is clearly proven. It is shown that

Table 6: Quantile regression estimators – Direct Carbon Emission

Decile	1	2	3	4	5	6	7	8	9
Coeff.	-0.004	-0.008	-0.008	-0.010	-0.012	-0.016	-0.021	-0.031	-0.055
p-value	0.127	0.007	0.004	0.000	0.000	0.000	0.000	0.000	0.000

Table 7: Quantile regression estimators – Indirect Carbon Emission

Decile	1	2	3	4	5	6	7	8	9
Coeff.	-0.087	-0.144	-0.164	-0.156	-0.162	-0.189	-0.220	-0.260	-0.409
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 8: Quantile regression estimators – Overall Carbon Emission

Decile	1	2	3	4	5	6	7	8	9
Coeff.	-0.004	-0.010	-0.008	-0.010	-0.010	-0.014	-0.019	-0.025	-0.045
p-value	0.081	0.007	0.008	0.000	0.000	0.000	0.000	0.000	0.000

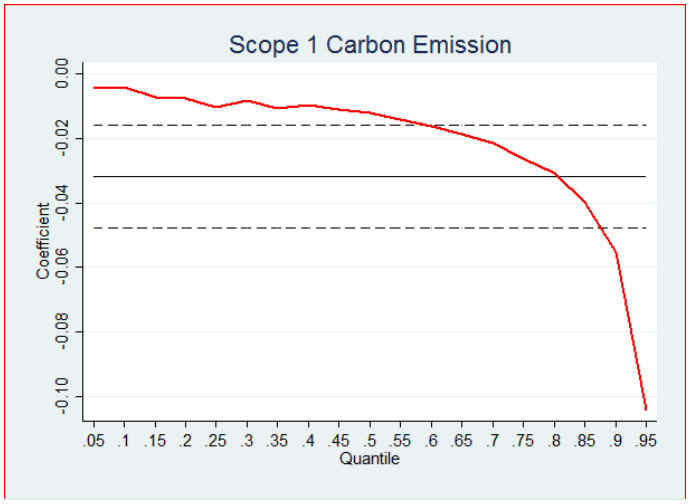


Figure 1: Quantile Regression Estimators - Direct Carbon Emission

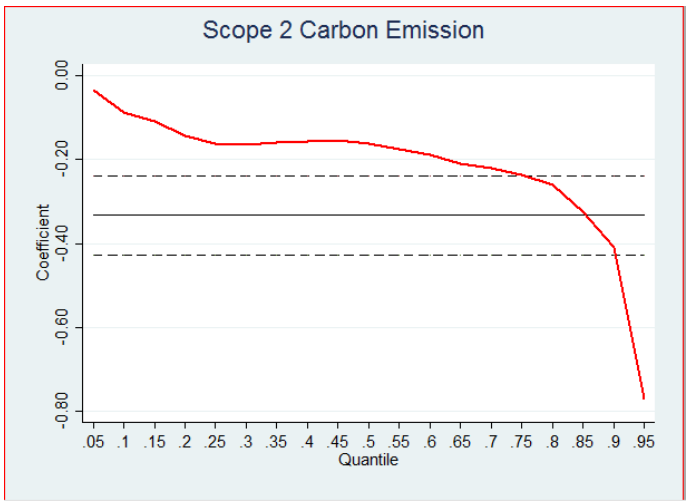


Figure 2: Quantile Regression Estimators - Indirect Carbon Emission

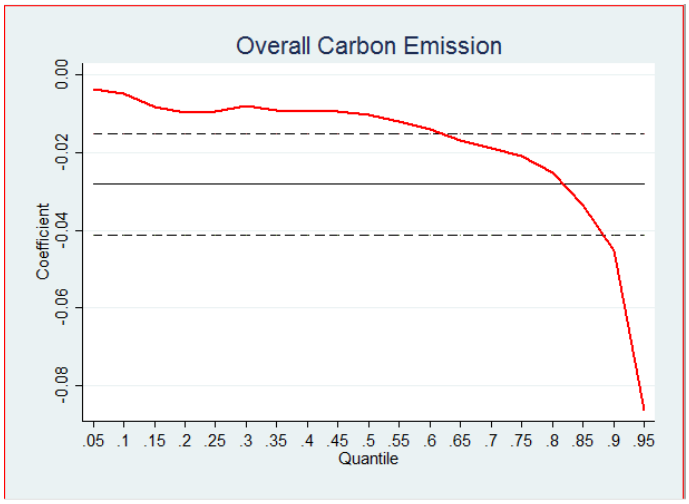


Figure 3: Quantile Regression Estimators - Overall Carbon Emission

the impact of carbon emissions on firm value depends significantly on the relative value of the respective firm. This indicates that the average effect is not informative and does not provide a reliable basis for taking adequate measures. However, as the relation between carbon emissions from all sources and firm value is estimated to be negative at the conditional mean as well as at the entire conditional distribution of Tobin's Q measurements, both approaches indicate that the market places high value to a firm's environmental performance.

The observed shift of the quantile estimators along the conditional distribution might be explained by the higher interest of investors in firms with higher Tobin's Q. As a high Tobin's Q indicates good management performance and returns that exceed the costs, such a firm is more attractive to potential investors (Lindenberg and Ross, 1981). Thus, the interest in the firm in general and therefore the interest in its environmental performance is higher. Consequently, one might conclude that, as well as it is likely that the effect of CSR on firm value depends on customer awareness (Servaes and Tamayo, 2013), the effect of carbon emissions on firm value depends on investor awareness and is therefore higher for firms with relatively higher Tobin's Q results.

6.2. Implications of the findings

The main result of the empirical analysis, suggesting that the effect of carbon emissions on firm value is significantly heterogenous, should be considered at all levels affected by the inevitable firm level carbon mitigation. In the following, some considerations on how the findings of this study may be taken into account at different levels are presented.

It has been shown that for a firm that is interested in increasing its value, carbon emission reduction is an adequate measure regardless of the value of the respective firm. This alone is an important finding for corporate managers who are hesitant about engaging in emission mitigation activities as they are sceptical about its effect on firm value. However, when referring to the average value-enhancing effect, an incorrect impression is given, because, for most firms, the real effect is either over- or underestimated. Hence, when evaluating to what extend firm value can be increased with carbon emission reduction and when comparing the value-enhancing effect to other 'value-increasing measures', the quantile estimators provide a more reliable base. Moreover, investors might also benefit from taking the quantile regression results into consideration. When being aware that a firm's engagement in emission mitigation activities does not increase firm value equally along the entire conditional distribution of firm value measurements, investment decisions can be designed accordingly and future profits can be estimated more reliable. Finally, the results might improve the efficiency of measures taken by regulators in order to encourage companies to reduce carbon emissions. By taking quantile effects into consideration, a focus can be set on firms that do not yet benefit from voluntary carbon emission reduction as much as others, namely relatively less valuable firms.

6.3. Limitations and suggestions for further research

A major limitation of this study is the application of a cross-sectional quantile regression model instead of a quantile regression model for panel data. As mentioned above, this is mainly due to the limited amount of time series in the present data. The cross-sectional approach reduces the informativeness of the results as it does not control for individual-specific unobserved heterogeneity (Huang et al., 2017). Further research might address this issue by exploiting data sets with a larger amount of time series or by applying improved quantile regression models for panel data.

Another question that should be answered by further research is why the value-enhancing of carbon emission mitigation is higher for relatively more valuable firms. As this study only estimated the nature of the relationship but did not analyse what causes the effects to increase with rising firm value, such a study would strongly improve the understanding of the connection between carbon emissions and firm value.

In addition, future research might also analyse and compare the long-term effects of carbon mitigation activities on firm value. Finally, an issue that might be investigated in further research is whether the effect of carbon emissions on other financial performance indicators than firm value is heterogenous as well. Especially corporate managers might be interested in the quantile-specific effects of carbon emissions on accounting-based measures.

7. Summary

In past decades, carbon dioxide emission has become a major environmental concern. To limit global warming and the associated negative consequences, the amount of carbon dioxide emitted from firms must be reduced (United Nations, 2015), because the corporate sector is responsible for significant quantities of overall carbon emissions (Aggarwal and Dow, 2011). To take appropriate measures, policy makers, investors and corporate managers have to deal with how the firm level mitigation of carbon emissions affects a firm's value (Lee and Min, 2015). In order to provide more reliable estimators on this relationship, a quantile regression approach was applied. In contrast to mean regression models, this approach addresses potential heterogenous effects and provides a more complete picture (Koenker and Hallock, 2001) of the relationship between carbon emissions and firm value. To improve the informativeness of the results, the effect of a firm's direct, indirect and overall carbon emissions on firm value was estimated separately. The estimation outcomes show that lower carbon emissions from all sources are significantly associated with higher Tobin's Q results. However, the quantile regression estimators revealed significant heterogeneity of the effects. It was found that a firm with relatively high firm value can benefit much more from reducing carbon emissions than a firm below the median of the conditional distribution of firm value measurement. By taking the results of this study into consideration, all levels affected by firm level emission mitigation might take more appropriate measures and improve the efficiency of those.

References

Aggarwal, R. and Dow, S. Greenhouse gas emissions mitigation and firm value: a study of large north-american and european firms. In *Midwest Finance Association 2012 Annual Meetings Paper*, 2011.

Al-Najjar, B. and Anfimiadou, A. Environmental policies and firm value. *Business Strategy and the Environment*, 21(1):49–59, 2012.

Bartram, S. M., Brown, G. W., and Conrad, J. The effects of derivatives on firm risk and value. *Journal of Financial and Quantitative Analysis*, 46(4): 967–999, 2011.

Capstick, S., Whitmarsh, L., Poortinga, W., Pidgeon, N., and Upham, P. International trends in public perceptions of climate change over the past quarter century. *Wiley Interdisciplinary Reviews: Climate Change*, 6(1): 35–61, 2015.

CDP Worldwide. Guidance for companies reporting on climate change on behalf of investors & supply chain members 2017 CDP. CDP Worldwide, 2017.

Cox, N. Winsor: Stata module to winsorize a variable. *Boston College Department of Economics*, 1998. URL https://ideas.repec.org/c/boc/bocode/s361402.html.

Dowell, G., Hart, S., and Yeung, B. Do corporate global environmental standards create or destroy market value? *Management Science*, 46(8):1059–1074, 2000.

Du, J., Wu, F., and Liang, X. Corporate liquidity and firm value: evidence from china's listed firms. In *EDP Sciences*, volume 24, pages 10–13. EDP Sciences, 2016.

Ganda, F. and Milondzo, K. The impact of carbon emissions on corporate financial performance: Evidence from the south african firms. *Sustainability*, 10(7):2398, 2018.

Gupta, K., Banerjee, R., and Onur, I. The effects of R&D and competition on firm value: International evidence. *International Review of Economics & Finance*, 51:391–404, 2017.

Huang, Q., Zhang, H., Chen, J., and He, M. Quantile regression models and their applications: a review. *J Biom Biostat*, 8:354, 2017.

IPCC. Summary for Policymakers of IPCC Special Report on Global Warming. IPCC Secretariat, 2018.

Iskandar, E. F. The Effect of Carbon Emissions Disclosure and Corporate Social Responsibility on the Firm Value with Environmental Performance as Variable Control. *Research Journal of Finance and Accounting*, 7(9): 122–130, 2016.

King, A. A. and Lenox, M. J. Does it really pay to be green? an empirical study of firm environmental and financial performance: An empirical study of firm environmental and financial performance. *Journal of Industrial Ecology*, 5(1):105–116, 2008.

Koenker, R. Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, 91(1):74–89, 2004.

Koenker, R. and Bassett Jr, G. Regression quantiles. *Econometrica: Journal of the Econometric Society*, pages 33–50, 1978.

Koenker, R. and Hallock, K. F. Quantile regression. *Journal of Economic Perspectives*, 15(4):143–156, 2001.

Konar, S. and Cohen, M. A. Does the market value environmental performance? *Review of economics and statistics*, 83(2):281–289, 2001.

Lee, K.-H. and Min, B. Green r&d for eco-innovation and its impact on carbon emissions and firm performance. *Journal of Cleaner Production*, 108:534–542, 2015.

Lindenberg, E. B. and Ross, S. A. Tobin's q ratio and industrial organization. *Journal of business*, pages 1–32, 1981.

Lindsey, R. Climate Change: Atmospheric Carbon Dioxide, 2018. URL https://www.climate.gov/news-features/understanding-climate/climate-change-atmospheric-carbon-dioxide. checked on 9/8/2018.

Matsumura, E. M., Prakash, R., and Vera-Muñoz, S. C. Firm-value effects of carbon emissions and carbon disclosures. *The Accounting Review*, 89(2): 695–724, 2014.

Mosteller, F. and Tukey, J. W. Data analysis and regression: a second course in statistics. *Addison-Wesley Series in Behavioral Science: Quantitative Methods*, 1977.

Nishitani, K. and Kokubu, K. Why does the reduction of greenhouse gas emissions enhance firm value? the case of japanese manufacturing firms. *Business Strategy and the Environment*, 21(8):517–529, 2012.

Powell, D. Quantile regression with nonadditive fixed effects. *RAND Cooperation*, 2015.

Powell, D. Quantile regression with nonadditive fixed effects. *Quantile Treatment Effects*, pages 1–28, 2016.

Schulze, N. Applied quantile regression: microeconometric, financial, and environmental analyses. 2004.

Servaes, H. and Tamayo, A. The impact of corporate social responsibility on firm value: The role of customer awareness. *Management Science*, 59(5): 1045–1061, 2013.

Sprinkle, G. B. and Maines, L. A. The benefits and costs of corporate social responsibility. *Business Horizons*, 53(5):445–453, 2010.

Union Investment. Wesentlichkeit von CO2-Emissionen für Investitionsentscheidungen. Union Investment, 2016.

United Nations. Paris Agreement. United Nations, 2015.

United States Environmental Protection Agency. International Climate Impacts, 2017. URL https://19january2017snapshot.epa.gov/climate-impacts/international-climate-impacts_.htm. checked on 9/8/2018.

Vermunt, R. Tobin's Q Ratio. What is and where can I find it?, 2013. URL http://finabase.blogspot.com/2013/05/tobins-q-ratio-what-is-and-where-can-i.html. checked on 9/26/2018.

Williams, R. G., Roussenov, V., Goodwin, P., Resplandy, L., and Bopp, L. Sensitivity of global warming to carbon emissions: Effects of heat and carbon uptake in a suite of earth system models. *Journal of Climate*, 30(23): 9343–9363, 2017.

World Resources Institute. A Corporate Accounting and Reporting Standard. The Greenhouse Gas Protocol, 2014.

Junior Management Science 4(3) (2019) 433-459

Junior Management Science

journal homepage: www.jums.academy

Die Legitimation einer Innovation durch Cultural Entrepreneurship – Explorative Fallstudie eines symbiotischen Zusammenspiels zwischen einem Start-up und dessen Schlüsselkunden

Leona Schink

Leuphana Universität Lüneburg

Abstract

Die „Cultural Entrepreneurship“ Literatur untersucht wie Start-ups die „Liability of newness“ durch kulturelle Überzeugungsleistungen überwinden, wobei meist Entrepreneure in der aktiven und die Bezugsgruppen (z.B. InvestorInnen) in der passiven Empfängerrolle gesehen werden. Die Fallstudie eines vom Start-up Tonio und dessen B-to-B Schlüsselkunden Turner Broadcasting System (TBS) durchgeführten und auf Tonios innovativer Technologie basierenden Pilotprojektes untersucht, wie Tonio sich bei TBS legitimiert und vor allem, wie interne Dynamiken innerhalb des Schlüsselkunden den Legitimationsprozess beeinflussen. Eine qualitative Inhaltsanalyse von Leitfadeninterviews, unternehmensinternen Dokumenten sowie frei zugänglichen Medienartikeln legt offen, dass sowohl Tonio, als auch interne Allianzpartner innerhalb von TBS in einer symbiotischen Zusammenarbeit aktiv für die Legitimation der innovativen Technologie gezielte kulturelle Überzeugungsarbeit leisten. So nutzt TBS bei der an interne Stakeholder gerichteten Überzeugungsarbeit beispielsweise Identitätsmechanismen, wie symbolische Handlungen, Narrative und Analogien, die in der vorherigen Phase ebenso von Tonio genutzt wurden, um TBS als Kunden zu gewinnen. Anschließend übernimmt Tonio Legitimationsarbeit von TBS zur Akquise neuer Kunden. Mit diesen Ergebnissen trägt diese Studie zum Verständnis von Innovationsprojekten als über Organisationsgrenzen hinweg fungierende Symbiosen bei.

Keywords: cultural entrepreneurship; legitimacy; new ventures.

1. Einleitung

Neue Organisationen, welche Innovationen am Markt etablieren wollen, leiden unter der „liability of newness“, einem Mangel an Legitimität oder sozialer Akzeptanz in der Wahrnehmung externer Stakeholder am Markt (Aldrich und Fiol, 1994; Zimmerman und Zeitz, 2002). Die Stakeholder sind sich unsicher, ob die neue Organisation sich angemessen und wünschenswert verhält, da noch keine Erfolgsbilanz zur Verfügung steht, um die Organisation zu bewerten (Aldrich und Fiol, 1994; Lounsbury und Glynn, 2001; Zott und Huy, 2007). Zur Überwindung dieser Unsicherheit und um externe Ressourcenhalter davon zu überzeugen, dass die neue Organisation dennoch Unterstützung verdient, können „Cultural Entrepreneurship“ Leistungen die Wahrnehmung der Stakeholder beeinflussen, indem deren Verständnis davon, was die neue Organisation ausmacht und was sie erreichen will, gelenkt wird (Fisher et al., 2017; Lounsbury und Glynn, 2001; Zimmerman und Zeitz, 2002; Zott und Huy, 2007). Bisher hat die „Cultural Entrepreneurship“ Literatur in diesem Zusammenhang die Schaffung einer Unternehmensidentität, unter anderem durch den Einsatz von Narrativen, symbolischen Handlungen oder Zertifizierungen, untersucht (Aldrich und Fiol, 1994; Fisher et al., 2017; Jennings et al., 2015; Lounsbury und Glynn, 2001; Rao, 1994; Zott und Huy, 2007).

An dieser Stelle möchte ich mich bei all denjenigen bedanken, die mich während der Anfertigung dieser Bachelorarbeit unterstützt und motiviert haben. Mein besonderer Dank gilt meinem Erstbetreuer Herrn Prof. Dr. Gegenhuber. Durch ihn kam ich erstmalig mit der Theorie des Cultural Entrepreneurship in Berührung und er war es, der den Kontakt zu Bernhard Prokop von Turner Broadcasting System herstellte und mir damit den Großteil der Datenbasis meiner Arbeit zugänglich machte. Durch seine intensive Betreuung und die strukturierten Ratschläge unterstützte er mich bei der Erstellung dieser Arbeit von Anfang an enorm. Darüber hinaus danke ich meinen Interviewpartnern von Turner Broadcasting System, Marc Goodchild und Bernhard Prokop sowie Dr. Florian Novak und Dr. Wolfgang Neubert vom Start-up Tonio. Abschließend möchte ich mich bei Turner Broadcasting System für die Übernahme der Reisekosten nach London und bei meinen Freunden und meiner Familie für das Korrekturlesen bedanken.

DOI: https://doi.org/10.5282/jums/v4i3pp433-459

„Cultural Entrepreneurship“ fußt somit auf dem grundlegenden Verständnis, dass ein Akteur einen anderen überzeugen will (Lounsbury und Glynn, 2001; Zott und Huy, 2007). Dabei werden auf Grund des bestehenden „entrepreneur-centric focus“ (Lounsbury et al., 2019b, S. 5) meist Entrepreneure in einer aktiven Rolle (Santos und Eisenhardt, 2009) gesehen, während deren Bezugsgruppen, beispielsweise in Form von Kunden oder Investoren, eher eine passive Empfängerrolle zugeschrieben wird (Fisher et al., 2017; Gehman und Soublière, 2017; Lounsbury und Glynn, 2001). In der jüngsten Zeit hingegen wird dazu tendiert, nicht mehr nur einzelne Akteure als aktiven und nach Legitimität strebenden Part anzuerkennen, sondern kulturelle Legitimationsarbeit als gemeinsame Anstrengung mehrerer Akteure zu begreifen. Dies spiegelt sich aktuell im Konzept des „Distributed Entrepreneurships“ (Garud et al., 2014; Gegenhuber, T., 2017; Gegenhuber und Naderer, 2018; Lounsbury et al., 2019b) wider. So zeigen erste Erkenntnisse, dass „Co-creation“ (Garud et al., 2014, S. 1180) oder „interactive dialogic process[es]“ (Gegenhuber und Naderer, 2018, S. 28) zwischen Organisationen und externen Stakeholdern im Entrepreneurshipkontext zu Legitimitätssteigerungen führen können.

Dieser vielversprechende neue Zugang zu kultureller Legitimationsarbeit erscheint außerdem in Verbindung mit Innovation (Garud et al., 2013; Van de Ven, 1986) interessant, da so bisher weitestgehend ungenutzte Potenziale aus den Parallelen von „Cultural Entrepreneurship“ und Innovationsmanagement ausgeschöpft werden können. So versucht die gegenwärtige „Cultural Entrepreneruship“ Literatur eine Verbindung zwischen Kultur, Entrepreneurship und Innovation zu schaffen und ein Verständnis von Innovation als kulturellen Prozess zu etablieren (Gehman und Höllerer, 2018; Lounsbury et al., 2019b).

Dass die Berücksichtigung der kulturellen Determinante von Innovationen Entrepreneuren die Legitimationsarbeit erleichtern kann, zeigten bereits Hargadon und Douglas (2001). Die Autoren erklärten die Legitimation des innovativen elektrischen Lichtes durch dessen Innovationsdesign, welches die Innovation in die vertraute Umwelt der zu überzeugenden Bezugsgruppen einbettete. Darauf aufbauend fanden Fisher et al. (2017) heraus, dass verschiedene Bezugsgruppen jeweils anders auf die Legitimationsarbeit von Technologie Start-ups reagieren, weshalb die kulturelle Überzeugungsarbeit von Start-ups nach den Autoren je nach Bezugsgruppe variieren sollte. Dennoch befasste sich die „Cultural Entrepreneurship“ Literatur seit dessen Aufkommen vor zwanzig Jahren lediglich mit einigen bestimmten Bezugsgruppen (Fisher et al., 2017; Lounsbury und Glynn, 2001). Gerade für Technologie Start-ups (Fisher et al., 2017) jedoch ist die Bezugsgruppe der B-to-B Schlüsselkunden für den Ressourcenerhalt und weiteres Wachstum höchst relevant (Perez et al., 2013). Zudem trägt die Ausrichtung der kulturellen Überzeugungsarbeit auf die Bezugsgruppe nachgewiesenermaßen zur Steigerung der sozialen Akzeptanz von Innovationen bei (Fisher et al., 2017; Lounsbury und Glynn, 2001; Zott und Huy, 2007).

Diese Arbeit nimmt daher die zwei Perspektiven ein, dass B-to-B Schlüsselkunden eine wesentliche Bezugsgruppe für „Cultural Entrepreneurship“ Leistungen darstellen, wobei vor allem aber die Perspektive, dass „Cultural Entrepreneurship“ eine Leistung mehrerer Akteure darstellt, als Motivation der Arbeit diente. Zusammenfassend bezieht sich das Interesse daher darauf, wie ein Start-up einen B-to-B Schlüsselkunden überzeugt und mobilisiert mit dessen innovativer Technologie zu arbeiten und wie interne Dynamiken innerhalb des Schlüsselkunden diesen Prozess beeinflussen.

Ein vielversprechender Schauplatz dafür ist das innovative Technologie Start-up Tonio aus Wien, welches das internationale Medienunternehmen Turner Broadcasting System (TBS) als B-to-B Schlüsselkunden gewann und daraufhin ein gemeinsames, auf Tonios innovativer Audio-Decoding-Technologie basierendes Pilotprojekt durchgeführt wurde. Eine Fallstudie dieser beiden Organisationen stellt dabei den empirischen Kontext dieser Arbeit dar (Edmondson und McManus, 2007; Yin, 2009). Als Analyseeinheit diente der Prozess der Gewinnung eines Schlüsselkunden für ein Pilotprojekt und die sich daraus ergebenden Konsequenzen, womit Tonio und TBS zusammen im Zentrum der Analyse standen. Hierfür waren Leitfadeninterviews mit Spezialisten aus beiden Organisationen die Hauptdatenquelle, während die theoretische Linse der „Cultural Entrepreneurship“ Literatur die Analyse antrieb.

Die Ergebnisse zeigen, dass sowohl Tonio als auch interne Allianzpartner innerhalb von dessen B-to-B Schlüsselkunden TBS im Laufe des Prozesses gemeinsam aktiv für die Legitimation der innovativen Technologie gezielte kulturelle Überzeugungsarbeit leisteten. Die „Cultural Entrepreneurship“ Literatur (Lounsbury und Glynn, 2001) und dabei insbesondere das Konzept des „Distributed Entrepreneurships“ (Garud et al., 2014; Gegenhuber und Naderer, 2018; Lounsbury et al., 2019b) sowie die Literatur zu Allianzen und Kooptation (Santos und Eisenhardt, 2009) werden durch die Charakterisierung von Innovationsprojekten, als über Organisationsgrenzen hinweg fungierenden Symbiosen, erweitert. Dies stellt den theoretischen Beitrag dieser Arbeit dar. Der Prozess wurde hierfür in drei Phasen mit verschiedenen Akteurs- und Bezugsgruppen-konstellationen sowie Legitimationsmechanismen aufgeteilt, wodurch sich die symbiotische Zusammenarbeit der Organisationen in drei Symbioseebenen zeigt. So nutzte TBS bei der an interne Stakeholder gerichteten Überzeugungsarbeit beispielsweise Identitätsmechanismen wie symbolische Handlungen, Narrative und Analogien, die zuvor in derselben Form von Tonio genutzt wurden, um TBS selbst zunächst als Kunden zu gewinnen. Außerdem ergibt die Analyse, dass auch Tonio in der anschließenden Phase als Teil der Symbiose Legitimationsarbeit von TBS übernahm, um diese im Zuge eines weiteren Wachstums bei der Akquisition neuer Kunden einzusetzen.

Neben dem theoretischen Beitrag besteht der empirische Beitrag dieser Arbeit in der Bestätigung bestehender Erkenntnisse zu „Cultural Entrepreneurship“ (Lounsbury und Glynn, 2001), symbolischem Management (Zott und Huy, 2007) und Innovationsdesign (Hargadon und Douglas, 2001) so-

wie der Präzision von Legitimationsarbeit für kompetenzsteigernde Innovationen (Abernathy und Clark, 1985; Tushman und Anderson, 1986) und relevanten Bezugsgruppen durch B-to-B Schlüsselkunden (Fisher et al., 2017).

2. Theoretischer Rahmen

2.1. „Cultural Entrepreneurship" – Eine Einordnung

Legitimität meint, dass eine Organisation hinsichtlich ihrer Normen, Werte und Handlungen von Stakeholdern in einem bestimmten sozialen Kontext als angemessen und erstrebenswert wahrgenommen wird (Suchman, 1995; in Fisher et al., 2017). Innovativen Start-ups fehlt es meist an Legitimität, da sie oft noch keinen „proven track record[s]" (Lounsbury und Glynn, 2001, S. 546) vorweisen können. Sie sind im Zuge ihres Wachstums dennoch auf die Unterstützung externer Akteure angewiesen (Fisher et al., 2017; Zott und Huy, 2007), was neue Organisationen vor eine kritische Herausforderung stellt (Aldrich und Fiol, 1994; Lounsbury und Glynn, 2001; Zott und Huy, 2007).

Die Frage danach, wie Start-ups trotz dieser fehlenden Legitimität am Markt Fuß fassen können, wurde durch das von Lounsbury und Glynn (2001) aufgestellte Konzept des „Cultural Entrepreneuships" erstmalig auf einer kulturellen Ebene verortet und dadurch ein „more fully contextualized understanding of entrepreneurship as socially and culturally embedded" (Lounsbury et al., 2019a, S. 4) gefördert. Von diesem Standpunkt aus muss die Legitimation einer Innovation durch kulturelle Managementarbeit erfolgen. Dadurch wird die Wahrnehmung der Innovation und der dazugehörigen neuen Organisation in den Augen der externen Stakeholder durch die Schaffung einer Unternehmensidentität beeinflusst und letztlich die Unsicherheit reduziert (Fisher et al., 2017; Lounsbury und Glynn, 2001). Storytelling ist einer von vielen zu „Cultural Entrepreneurship" zählenden Mechanismen, bei welchem Start-ups Stakeholdern verständlich machen, was ihre Organisation ausmacht und wieso sie Unterstützung verdient (Aldrich und Fiol, 1994; Lounsbury und Glynn, 2001). Hierbei wird sowohl die Organisation selbst als auch das breitere institutionelle Umfeld durch die Beschreibung des firmenspezifischen und des institutionellen Kapitals dargestellt, wobei im Zuge einer optimalen Unverwechselbarkeit die Balance zwischen Gleichheit und Einzigartigkeit gefunden werden soll (Lounsbury und Glynn, 2001).

Weitere Beispiele für „Cultural Entrepreneurship" Anstrengungen sind symbolische Handlungen (Aldrich und Fiol, 1994; Jennings et al., 2015; Zott und Huy, 2007) oder symbolische Aktionen wie Zertifizierungen (Rao, 1994). Symbolische Handlungen können auf mehreren Wegen die Glaubwürdigkeit von Organisationen steigern. Funktionierende Produkte oder Technologien wie Prototypen übermitteln potenziellen Ressourcengebern so beispielweise organisationale Erfolge und schaffen Vertrauen (Zott und Huy, 2007). Von Entrepreneuren genutzte „Symbolic Devices" linguistischer, visueller oder performativer Art, wie zum Beispiel tiefgreifende Fragen, Bilder oder Analogien, können zudem Emotionen in Stakeholdern hervorrufen, wodurch diese sich mit neuen Projekten identifizieren (Jennings et al., 2015).

In diesem Sinne werden Entrepreneure als „skilled cultural operators who shape interpretations of the nature and potential of their new venture to those who may supply needed resources" (Lounsbury und Glynn, 2001, S. 549) gesehen, welche Kultur als „set of resources" oder „toolkit" (Lounsbury et al., 2019a, S. 5) in der Legitimationsarbeit einsetzen. Auch im Zuge der Forschung an der Einführung des innovativen elektrischen Lichtes durch Thomas Edison wurde die kulturelle Determinante einer Innovation im Zusammenhang mit Entrepreneurship untersucht (Hargadon und Douglas, 2001). Dabei gingen die Autoren der Frage nach, wie Innovationen Einzelpersonen und Organisationen verständlich gemacht werden können und kamen unter anderem zu dem Ergebnis, dass das Design von Innovationen zwischen diesen und etablierten Institutionen in einem Sozialsystem vermitteln kann. Thomas Edison fügte dem elektrischen Licht bestimmte Elemente („Skeumorphs"), wie die Zählmethode der Vorgängertechnologie hinzu, welche objektiv keine funktionale Absicht besaßen und allein dazu dienten, das Verständnis der Öffentlichkeit anzuregen. Neben Lounsbury und Glynn (2001) zeichnen daher auch Hargadon und Douglas (2001) ein kulturelles Bild von Innovation und Entrepreneurship, in welchem Entrepreneure als aktive kulturelle Manager agieren.

Dass aufstrebende Organisationen und deren Innovationen durch fehlende Legitimität vor ein kritisches Problem gestellt werden und wie die Fachliteratur des „Cultural Entrepreneur-ship" dieses unter anderem durch Narrative und symbolisches Management zu lösen versucht, wurde nun verdeutlicht. Auf diesem Verständnis wird aufgebaut, indem die nachfolgend unter den Punkten 2.2 und 2.3 beschriebenen zwei Perspektiven eingenommen werden. Hierdurch wird zum Erkenntnisgewinn über die kulturellen Triebkräfte von Entrepreneurship und Innovation beigetragen, welche bisher in der Wissenschaft durch den starken Fokus auf dahinterstehende wirtschaftliche Dynamiken wenig Beachtung fanden (Lounsbury et al., 2019a).

2.2. „Cultural Entrepreneurship" als gemeinsame Leistung mehrerer Akteure

Bezüglich der Akteure, welche die kulturellen Überzeugungsleistungen erbringen, existiert momentan ein „entrepreneur-centric focus in much entrepreneurship research" (Lounsbury et al., 2019b, S. 5), denn abgesehen von Start-ups wurde sich bisher wenig mit anderen Erbringern von kultureller Managementarbeit auseinandergesetzt. Zu den wenigen Erkenntnissen hierzu zählt die Forschung von Gegenhuber und Naderer (2018), die „Cultural Entrepreneurship" als dialogischen Prozess begreift, in welchem Entrepreneure und die Träger ihrer kulturellen Überzeugungs-leistungen aktiv miteinander interagieren. Die Autoren fanden durch die Untersuchung von Organisationen im Bereich Gaming und deren Communities heraus, dass diese sich im Dialog auf bestimmte geltende Verhaltensnormen einigten und der

Umgang mit diesen Normen durch die Organisation wiederum die Legitimität beeinflusste. Auch Garud et al. (2014) beschäftigten sich im Entrepreneurshipkontext sowohl mit Innovation als Prozess als auch damit, wie Innovation durch verschiedene Akteure und Kontexte „co-created" (Garud et al., 2014, S. 1180) wird. Dabei gehen sie insbesondere auf Narrative ein, mit welchen Entrepreneure Innovationen kontextualisieren:

> „Entrepreneurs recombine elements from different contexts to articulate and implement opportunities that are exciting to themselves and to stakeholders including investors, regulators, and advisors. When presented with entrepreneurial narratives, stakeholders actively read into them from their own perspectives (Ricoeur, 1984). This, in turn, generates fresh interpretations, some of which entrepreneurs incorporate into their narratives. Consequently the narratives that emerge through such interactions and dialogue reflect not just the aspirations and motivations of the entrepreneurs, but also bear an imprint of the inputs that stakeholders offered" (Garud et al., 2014, S. 1182).

Hier wird deutlich, dass nicht nur Entrepreneure, sondern auch Stakeholder einen Einfluss auf Narrative und somit auf Legitimationsarbeit im Sinne von „Cultural Entrepreneurship" haben können. Doch die Ergebnisse von Garud et al. (2014) liefern bezüglich der Effekte von Bezugsgruppen auf die Narrative der Entrepreneure keine konkreten Erkenntnisse. Trotzdem erweitern die Autoren, genauso wie Gegenhuber und Naderer (2018), mit ihren jüngsten Erkenntnissen das simple Kommunikationsmodell von Organisationen als Sender an passive Empfänger, welche diese Bemühungen lediglich für sich auswerten (Gehman und Soublière, 2017; Lounsbury und Glynn, 2001). In Übereinstimmung damit wurde herausgefunden, dass Legitimitätssteigerungen auch erreicht werden können, wenn Entrepreneure als Erbringer kultureller Leistungen an bestimmte Bezugsgruppen gemeinsam mit diesen (proaktiv) im Rahmen von „Distributed Entrepreneurship" kulturelle Managementarbeit leisten (Garud et al., 2014; Gegenhuber, T., 2017; Gegenhuber und Naderer, 2018; Lounsbury et al., 2019b).

Das Verständnis von „Cultural Entrepreneurship" als gemeinsame Bemühung im Kontext der Legitimation von Innovationen erscheint dabei insbesondere im Zusammenhang mit B-to-B Schlüsselkunden relevant. Diesbezüglich wurde sich in der Innovationsforschung bereits ausführlich mit den Herausforderungen eines erfolgreichen Innovationsmanagements innerhalb reifer Organisationen beschäftigt (z.B. Garud et al., 2013; Van de Ven, 1986). Wenn durch eine Innovation also beispielsweise die Kompetenz einer etablierten Organisation verstärkt werden soll (Abernathy und Clark, 1985; Tushman und Anderson, 1986), stellt eine solche Umstellung reife Organisationen häufig vor Probleme (Van de Ven, 1986) oder Komplexitäten (Garud et al., 2013). Das Innovationsmanagement wird nach Van de Ven (1986) unter anderem damit konfrontiert, dass die Mitarbeiter oftmals lieber an bestehenden Praktiken festhalten, als sich neuen Ideen oder Notwendigkeiten zu öffnen, was zudem umso schwieriger wird, desto erfolgreicher die Organisation ist. Darüber hinaus ist die Implementierung von Innovationen ein „collective achievement of pushing and riding [those] ideas into good currency" (Van de Ven, 1986, S. 591), welches häufig von politischen Dynamiken überschattet wird. Hinzu kommt, dass Individuen auf Grund der diversen eingesetzten Funktionen und Ressourcen den Überblick verlieren können. Eine essenzielle Rolle spielt daher der General Manager, welcher eine Infrastruktur schaffen kann, um die erforderliche Umstrukturierung bestmöglich zu leiten (Van de Ven, 1986).

Für sich im Innovationskontext bewegende General Manager innerhalb von B-to-B Schlüsselkunden ist es demnach selbstverständlich, leitend aktiv zu sein und komplexen Problematiken durch die Überwindung interner Barrieren zu begegnen. Angesichts dieser Kompetenz erscheint es vielversprechend zu untersuchen, inwiefern gerade General Manager innerhalb von B-to-B Schlüsselkunden auf Grund ihrer Erfahrung und Funktion in der Organisation fähig sind, sich gemeinsam mit den innovierenden Entrepreneuren intern für die Überwindung der Unsicherheit bezüglich der Innovation einzusetzen. Darum soll in dieser Arbeit auch untersucht werden, ob und wie die Bezugsgruppe der Schlüsselkunden nach der Überzeugung durch das Start-up nicht selbst zum kulturellen Akteur wird und so im Sinne von „Distributed Entrepreneurship" gemeinsam mit dem Start-up im Legitimationsprozess agiert. Eine wissenschaftliche Evidenz für eine solche Teil-Entrepreneurship-Rolle würde dem aktuellen Forschungsaufruf aus der Wissenschaftsgemeinschaft des „Cultural Entrepreneurships" gerecht werden und so die bisher vernachlässigte, aber durchaus zentrale Rolle von Kultur in den Bereichen Entrepreneurship und Innovation belegen (Gegenhuber und Naderer, 2018; Lounsbury et al., 2019a).

2.3. B-to-B Schlüsselkunden: Wesentliche Bezugsgruppe für kulturelle Überzeugungsarbeit

Durch die Einnahme einer zweiten Perspektive sollen die existierenden Bezugsgruppen von „Cultural Entrepreneurship" Leistungen erweitert und genauer untersucht werden. Lounsbury und Glynn (2001) nennen Wettbewerber, Kunden, Investoren und andere relevante Stakeholder als potenzielle Empfänger von Storytelling, doch in ihren Beispielen versuchen neue Organisationen hauptsächlich Investoren von sich zu überzeugen. Wie andere Bezugsgruppen beeinflusst werden können, ob und welche weiteren Bezugsgruppen existieren, wird kaum thematisiert. Auch Hargadon und Douglas (2001) untersuchten lediglich, wie Neues in die vertraute Umwelt von Einzelpersonen und der breiten Öffentlichkeit eingebettet werden kann.

Fisher et al. (2017) befassten sich bei ihrer Forschung zu Legitimitätsmechanismen differenzierter mit Geldgebern und fanden heraus, dass diese insofern heterogen sind, als dass bei diesen jeweils andere Beweggründe für Investitionsentscheidungen zur Unterstützung neuer Organisationen

existieren. Die Autoren unterteilen Legitimitätsmechanismen in die Kategorien Identität, Assoziation und Organisation, auf welche Geldgeber durch individuelle Logiken anders reagieren. Beispielsweise legitimieren sich junge technologiebasierte Organisationen in den Augen von „Angel Investors" im Rahmen von Identitätsmechanismen durch eine disruptive Rhetorik, während Unternehmen bei Crowdfunding mit einer Community-Rhetorik punkten können (Fisher et al., 2017). Dass sich die Akteure innerhalb der Bezugsgruppen von „Cultural Entrepre-neurship" Leistungen nach der Art und Weise unterscheiden, wie Entrepreneure sich bei ihnen legitimieren können, ist besonders im Innovationskontext relevant, da Innovationen für die zu überzeugenden Einheiten etwas gänzlich Neues darstellen und somit eine besonders hohe Unsicherheit beinhalten (Aldrich und Fiol, 1994).

Überträgt man diese Erkenntnis auf „Cultural Entrepreneurship", dann benötigen innovierende Entrepreneure detailliertes Wissen darüber, wie sie durch das bewusste Setzen von Akzenten ihr Storytelling oder ihre symbolischen Handlungen gezielt an die verschiedenen Bezugsgruppen anpassen können, um die Ressource der Legitimität für sich und ihre Technologie zu gewinnen (Zimmerman und Zeitz, 2002). Die Ergebnisse von Fisher et al. (2017) sind jedoch insofern begrenzt, als dass sie zwar differenzierte Erkenntnisse zu Überzeugungsmöglichkeiten verschiedener Bezugsgruppen liefern, diese sich aber lediglich auf wenige bestimmte beziehen („Crowdfunding backers, government agencies, angel investors, venture capitalists, corporate venture capitalists"). Doch da die Autoren wiederum gerade technologiebasierte Start-ups in der Rolle der Überzeuger sehen, ist es umso erstaunlicher, dass B-to-B Schlüsselkunden nicht zu den wichtigsten Bezugsgruppen gezählt werden, da diese sowohl die Rolle des Investors als auch die des besten und wichtigsten Kunden für neue technologiebasierte Organisationen einnehmen können (Perez et al., 2013). Daher wird in dieser Arbeit neben der Perspektive auf „Cultural Entrepreneurship" als gemeinsame Leistung mehrerer Akteure zusätzlich die Perspektive eingenommen, dass vor allem B-to-B Schlüsselkunden zu den wichtigsten Bezugsgruppen von „Cultural Entrepreneurship" Arbeit durch neue Organisationen zählen. Durch diese Sichtweise wird der aktuelle Fokus in der einschlägigen Literatur von den Bezugsgruppen der Geldgeber (Fisher et al., 2017) und der Medien (Petkova et al., 2013) auf die der B-to-B Schlüsselkunden verlagert.

Da sich die „Cultural Entrepreneurship" Literatur bisher vorrangig auf die Sender, in Form von neuen Organisationen, konzentrierte (Lounsbury und Glynn, 2001; Zott und Huy, 2007) und die Bezugsgruppen somit noch nicht ausreichend untersucht wurden, liegen außerdem noch keine Erkenntnisse dazu vor, wie die Bezugsgruppe der B-to-B Schlüsselkunden bestmöglich überzeugt werden kann. Dieses Wissen könnte für innovierende Start-ups jedoch eine kritische Möglichkeit der Kundenakquise im Bereich B-to-B darstellen.

Ebenso kritisch wäre für neue Organisationen das Wissen darüber, wie im Zuge einer erfolgreichen Überzeugung von potenziellen B-to-B Schlüsselkunden die zu schaffenden Unternehmensidentitäten bestmöglich an den Charakter der Innovation angepasst werden können. Dieser zeichnet sich nach Tushman und Anderson (1986) entweder in der Steigerung oder der Zerstörung der Kompetenz etablierter Organisationen aus. Im Zusammenhang mit „Cultural Entrepreneurship" wurde die Idee der kompetenzsteigernden und kompetenzzerstörenden Innovationen von Lounsbury und Glynn (2001) zwar aufgegriffen, jedoch nur oberflächlich und einseitig behandelt. Die Autoren geben lediglich ein Beispiel einer kompetenzzerstörenden Innovation und gelangen dennoch zu der Einsicht, dass Organisationen mit kompetenzsteigernden Innovationen Unternehmensidentitäten schaffen sollten, welche betonen „how they fit into or complement extant industry practices" (Lounsbury und Glynn, 2001, S. 558), ohne die Frage nach dem „wie" zu beantworten.

Zusammenfassend ist es für innovative Organisationen also von höchster Wichtigkeit, dass Neues für Bezugsgruppen verständlich gemacht wird, wodurch Unsicherheiten abgebaut, Legitimität gewonnen und letztlich Kooperationen eingegangen werden können (Fisher et al., 2017; Gegenhuber und Naderer, 2018; Lounsbury und Glynn, 2001; Zimmerman und Zeitz, 2002; Zott und Huy, 2007). An den in diesem Zusammenhang relevanten Aspekten der Bezugsgruppen, insbesondere den B-to-B Schlüsselkunden und der Teilrolle dieser als kulturelle Akteure in einem ohnehin jungen wissenschaftlichen Feld, besteht jedoch noch Forschungsbedarf. Daher führen die zwei beschriebenen Perspektiven auf die Problematik der Legitimation von Innovationen durch „Cultural Entrepreneurship" zum Ausgangspunkt für die empirische Forschung, in welcher untersucht werden soll, wie eine neue Organisation B-to-B Schlüsselkunden überzeugt und mobilisiert mit dessen innovativer Technologie zu arbeiten und wie interne Dynamiken innerhalb des Schlüsselkunden diesen Prozess beeinflussen.

3. Methodik und empirischer Kontext

3.1. Forschungsdesign

Das Forschungsinteresse bestand in der Untersuchung von „Cultural Entrepreneurship" als gemeinsamer Leistung mehrerer Akteure sowie von B-to-B Schlüsselkunden als Bezugsgruppe dieser kulturellen, von innovativen Organisationen geleisteten Überzeugungsarbeit und den dabei wirkenden internen Dynamiken. Damit einhergehend wurde durch universitäre Kontakte der Zugang zu Spezialisten innerhalb des hochinteressanten und vielversprechenden innovativen Technologie Start-ups Tonio sowie zu dessen B-to-B Kunden TBS erleichtert. Aus diesen Gründen wurde sich für die Erstellung einer empirischen Arbeit und die Durchführung einer Fallstudie entschieden. Da im Zuge einer Fallstudie mit einer „Wie"-Frage ein komplexes, aktuelles und bisher wenig erforschtes Phänomen tiefgehend untersucht werden sollte, schien dieser Forschungsansatz geeignet zu sein (Edmondson und McManus, 2007; Yin, 2009).

Um das Phänomen auf Basis einer systematischen Analyse besser zu verstehen, stellten zudem qualitative Interviews mit Spezialisten aus der Praxis die passende Methode dar

(Kaiser, 2014). Diese wurden als Leitfadeninterviews durchgeführt, da Spezialisten auf Grund ihrer Funktion in ihrem Feld häufig unter Zeitdruck stehen und Leitfäden durch ihre Steuerungsfunktion den Ausschluss unergiebiger Themen erlaubten (Flick, 2007). Durch die Entscheidung für Leitfadeninterviews konnten „spezifische Informationen über ein zu untersuchendes Phänomen [generiert werden], die anderweitig nicht zu erhalten wären“ (Kaiser, 2014, S. 31). Als Analyseeinheit diente der Prozess der Gewinnung eines Schlüsselkunden für ein Pilotprojekt und dessen Konsequenz. Dementsprechend wurde sich auf die gemeinsame Aktivität von Tonio und TBS konzentriert, weshalb diese zwei Organisationen den breiteren empirischen Kontext darstellten.

3.2. Fallbeschreibung

Tonio (Ton mit Information) ist ein Wiener Technologie Start-up, welches eine innovative Audio-Decoding-Technologie entwickelte. Unter digitaler Innovation wird nach Nambisan et al. (2017) die „creation of (and consequent change in) market offerings, business processes, or models that result from the use of digital technology“ (Nambisan et al., 2017, S. 224) verstanden, weshalb nach dieser Definition auch Tonios digitale Technologie eine digitale Innovation darstellt. Durch einen unhörbaren Code, welcher beispielsweise in das Audiosignal einer Fernsehserie oder einer Live-Übertragung eingearbeitet wird, befähigt die Technologie Rundfunkunternehmen, wie TV- oder Radiosender, dessen Inhalte zu versenden und dieses Signal im nächsten Schritt auf Smartphones und Tablets zu decodieren und entsprechend darzustellen (EQ# 13 Tonio, 2018). Die Technologie aktiviert somit absolut synchron zur Ausstrahlung (EQ# 13 Tonio, 2018) den Second-Screen und ermöglicht dem Sender aktiv an den Empfänger heranzutreten. 2014 fand der Markteintritt statt und seitdem setzten Kunden wie das Medienunternehmen Cinecom oder der Österreichische Rundfunk Tonios neue Technologie ein. In Kooperation mit dem öffentlich-rechtlichen Medienanbieter entwickelte Tonio beispielsweise 2016 für die Zuschauer einer Kriminalserie ein Lügendetektor-Erlebnis am Smartphone. Hierbei wurden im Verlauf der Ausstrahlung der Episoden Hintergrundinformationen zu den Charakteren freigeschaltet und das Smartphone wurde zum Lügendetektor, welcher den Zuschauern ausstrahlungssynchron signalisierte, sobald ein Protagonist log (EQ# 14 Die Presse, 2018). Mit dieser innovativen Technologie versucht Tonio die Kompetenz etablierter Unternehmen in der Unterhaltungsindustrie zu stärken (vgl. Tushman und Anderson, 1986), was durch die Betrachtung von dessen großem Schlüsselkunden TBS und der aktuellen Entwicklung der Medienbranche deutlich wird.

Turner Broadcasting System ist ein internationales Medienunternehmen und neben HBO und Warner Bros. Entertainment eine Tochter der WarnerMedia. Als Produzent von Nachrichten- und Unterhaltungsprogrammen ist TBS Anbieter von Programminhalten für TV, Internet und mobile Plattformen. Neben CNN und Cartoon Network gehören zu TBS weit über 100 internationale Fernsehsender. Die Medienbranche, in welcher sich die Organisation befindet, erfährt aktuell durch die Digitalisierung und das Internet einschneidende Veränderungen. Digitale Technologien gestalten grundlegend neu, wie Inhalte produziert, verbreitet und konsumiert werden. Auf dem deutschen TV-Markt wurde das traditionelle Fernsehen, zu welchem auch TBS zählt, von Streaminganbietern wie Netflix und Amazon Prime bei unter 30-jährigen Zuschauern beispielsweise bereits überholt, während die TV-Gesamtumsätze (ohne Werbung) immer weiter sinken (EQ# 15 PwC, 2018; EQ# 16 Deloitte, 2018). Die Konsumenten verlangen, dass attraktive und relevante Inhalte jederzeit und überall abrufbar sind.

Der Schlüssel zur Zukunftsfähigkeit von heute etablierten Fernsehsendern liegt damit im Ausbau der Digitalkompetenz (EQ# 17 W&V, 2018). Um den digitalen Medien nicht zu unterliegen und den Konsumveränderungen Rechnung zu tragen, entwickelt auch TBS sein Geschäftsmodell in neu gegründeten Abteilungen wie „Digital Ventures and Innovation“ weiter, indem sich dort auch mit der Thematik der digitalen Transformation oder Direct-to-Consumer Konzepten auseinander-gesetzt wird. Weiterhin wurde die Mission formuliert, das Fernsehen neu zu erfinden und dessen Zuschauer zu loyalen Fans der Inhalte zu machen (EQ# 18 TBS, 2019). In diesem Zusammenhang soll auch die Kooperation mit Technologieanbietern wie Tonio die Digitali-sierung des Kerngeschäftes, dem linearen Fernsehen, schrittweise voranbringen.

3.3. Datensammlung

Die Datenbasis der Fallstudie bestand aus drei Datenquellen: Leitfadeninterviews, von TBS bereitgestellte, unternehmensinterne Dokumente, hauptsächlich bestehend aus PowerPoint Präsentationen sowie im Internet frei zugängliche Informationen von Websites oder Medienartikeln, welche dem Verständnis der Handlungen der beiden Organisationen als kulturellen Akteuren dienlich waren. Diese unterschiedlichen Datenquellen dienten einem möglichst ganzheitlichen Verständnis und ermöglichten Triangulation, wodurch kritische Punkte der Analyse innerhalb mehrerer Quellen belegt wurden (Lamnek und Krell, 2010). Die PowerPoint Präsentationen und Berichte sind auf Zeitpunkte von Ende 2016 bis Mitte 2018 datiert und dienten der internen Vorstellung von Tonio und dessen Audio-Decoding Technologie bei verschiedenen Stakeholdern innerhalb der Organisation, des Pilotprojektes und dessen Performance als auch der Darstellung von weiteren möglichen, auf der Tonio-Technologie basierenden Produkten und deren Umsetzung am Markt. Die Leitfadeninterviews jedoch stellten die Hauptdatenquelle für die Beleuchtung des Prozesses und der Konsequenzen der Überzeugung eines Schlüsselkunden durch ein innovatives Start-up im Sinne von „Cultural Entrepreneurship“ dar.

Eine intensive Auseinandersetzung mit den internen Materialien von TBS sowie eine Internetrecherche zu beiden Organisationen und zu den einzelnen Interviewpartnern ermöglichte zunächst ein Verständnis der jeweiligen Unternehmensgegenstände und damit die Einordnung des Falles in den aktuellen Branchenkontext. Danach wurden im November und Dezember 2018 insgesamt vier Leitfadeninterviews

Tabelle 1: Datenbasis der Fallstudie

Datentyp	No.	Beschreibung	Datum
Leitfadeninterviews	4	Bernhard Prokop (TBS) Marc Goodchild (TBS) Dr. Florian Novak (Tonio) Dr. Wolfgang Neubert (Tonio)	London, November 2018 London, November 2018 Wien, Dezember 2018 Wien, Dezember 2018
Interne Dokumente von TBS	8	6 PowerPoint Präsentationen / Berichte 1 Marktübersicht „Gumball VIP“ 1 Vorstellung Tonios an AT&T	Bereitgestellt im Oktober 2018 Entstehungsdaten zwischen Ende 2016 bis Mitte 2018
Medienartikel & Websites	16	Medienartikel, Websites, Newsletter Tonio	Oktober 2017 und Ende 2018

geführt: Zwei mit Interviewpartnern innerhalb Tonios und zwei innerhalb dessen Schlüsselkunden TBS. Durch Kontakte der Leuphana Universität Lüneburg konnte zunächst mit dem „Director of Digital Transformation“ von TBS in Kontakt getreten werden, welcher sehr eng mit dem CEO von Tonio zusammenarbeitet und den Kontakt zu diesem herstellte. Nach der ersten Kontaktaufnahme schlugen beide weitere Interviewpartner vor, sodass sich insgesamt vier geeignete und in unmittelbarem Zusammenhang mit der Technologie stehende Partner für die Interviews fanden.

Im Zuge der zielgerichteten Interviewvorbereitung wurden durch Telefon und E-Mail grundlegende Fragen zu den Aufgabenbereichen der Erzählpersonen gestellt. Um auch die Erzählpersonen bestmöglich auf das Interview vorzubereiten, wurde diesen im Vorfeld ein Informationsblatt zum Forschungsprojekt und dem Interviewablauf zugeschickt. Da sich alle vier Interviewpartner außerhalb von Deutschland befanden, wurde für die Interviews angereist. Bezüglich der Interviewsituation hatte dies den Vorteil, dass diese an den Arbeitsplätzen der Befragten durchgeführt wurden, wo sie sich wohl und sicher fühlten (Helfferich, 2011). Es wurde außerdem auf eine günstige Sitzanordnung geachtet und eine Anordnung der Stühle an einem Tisch über Eck gewählt (Helfferich, 2011).

Während der Interviewsituation wurde sich an das in der qualitativen Forschung zentrale Grundprinzip der Offenheit (Helfferich, 2011) gehalten. Dieses verlangt, dass der Erzählperson „Raum“ gegeben wird, ihr eigenes Relevanzsystem oder ihre Deutungsmuster zu entfalten und dieser keinesfalls das eigene Relevanzsystem überzustülpen. Auf der Ebene der Interviewsteuerung bedeutet Offenheit, die Erzählpersonen stets aussprechen zu lassen und ihnen Zeit für Pausen einzuräumen. Während der Gespräche wurde sich darauf fokussiert, die Entscheidungen und Ereignisse zu verstehen, welche die Befragten im Laufe der Kooperation für wichtig hielten. Die in diesem Bewusstsein geführten Interviews dauerten zwischen 49 Minuten und 120 Minuten. Die Interviews wurden aufgenommen und im Nachgang transkribiert. Im unmittelbaren Anschluss an die Interviews wurden diese mit einem Protokollbogen dokumentiert, welcher Fragen zur Gesprächsatmosphäre, zu Reaktionen des Gesprächspartners und zu besonderen personalen Beziehungen zwischen der Erzählperson und der interviewenden Person beinhaltete. Darüber hinaus wurden vor und nach den Interviews gemachte Aussagen der Gesprächspersonen durch handschriftliche Notizen festgehalten.

3.3.1. Vorstellung der Interviewpartner

Die Zielgruppe für Leitfadeninterviews sind „in der Regel Mitarbeiter einer Organisation in einer spezifischen Funktion mit einem bestimmten (professionellen) Erfahrungswissen“ (Flick, 2007). Dieses Wissen besteht nicht nur aus systematisiertem und reflexiv zugänglichem Fachwissen, sondern besitzt „zu großen Teilen den Charakter von Praxis- oder Handlungswissen [...]“ (Bogner und Menz, 2002; in Flick, 2007, S. 215). Obwohl persönliche Kontakte die Auswahl und den Zugang zu den Erzählpersonen erleichterten, begründete sich die Auswahl dieser allein aus der Fragestellung.

Dr. Florian Novak (CEO und Gründer von Tonio)

Florian Novak ist Medienunternehmer mit langjähriger Erfahrung in der Radio- und Rundfunkbranche. Neben seiner Arbeit bei Tonio ist er aktuell Geschäftsführer von LoungeFM, einem österreichischen Radiosender, mit welchem er im Jahr 2005 Schlagzeilen als „Erste[r] Handy-Radiosender Österreichs“ (EQ# 19 Radioszene, 2005) machte. Er ist seit 2017 promovierter Jurist der Universität Wien und befasste sich in seiner Dissertation mit dem öffentlich-rechtlichen Auftrag des Österreichischen Rundfunks.

Dr. Wolfgang Neubert (Senior Berater und Business Angel bei Tonio)

Wolfgang Neubert wurde von Florian Novak im Jahr 2016, zwei Jahre nach dem Markteintritt des Start-ups, als Strategiepartner ins Team geholt und war dadurch stets in wichtige Kundengespräche mit eingebunden. Neubert besitzt jahrzehntelange Erfahrung in der Strategieberatung bei McKinsey & Company und ist in unterschiedlichen Vorstandsfunktionen in österreichischen Aktiengesellschaften tätig (EQ# 20 Ashoka, 2016). Er ist promovierter Ingenieur und befasste sich in seinem Studium mit Nachrichtentechnik.

Bernhard Prokop (Director of Digital Transformation, Turner International)

Bernhard Prokop ist bei TBS in der Abteilung „Digital Venture & Innovation“ verantwortlich für die Identifikation und Aus-

führung neuer Initiativen, wozu unter anderem die Nutzung neuer Technologien gehört. In diesem Zusammenhang holte er auch Tonio und dessen innovative Technologie ins Unternehmen. Er hat einen Hintergrund im strategischen Management und war in seiner Vergangenheit als Unternehmensberater, Ministersekretär und im Bereich des Videojournalismus tätig, wonach er 2015 zu TBS nach London kam. Zudem studierte er Organisationstheorie an der Johannes Kepler Universität Linz auf Diplom.

Marc Goodchild (Head of Digital Content Strategy & Product for Kids, Turner EMEA)

Marc Goodchild ist als Kollege von Bernhard Prokop für die operative Implementierung und Distribution neuer Technologien innerhalb der verschiedenen Märkte von TBS verantwortlich. Bevor er 2017 zu TBS kam und dort im Rahmen des Pilotprojektes als Hauptverantwortlicher mit Tonios Technologie zu arbeiten begann, arbeitete er als Fernsehproduzent und „Digital Executive“ im Broadcast-, Produktions- und Start-up Bereich

3.3.2. Leitfadenerstellung

Vor der Sammlung von Fragen für die Leitfäden wurde sich mit der Erzählperspektive und der eigenen Erwartungshaltung auseinandergesetzt. Dies verfestigte die Bereitschaft, in den Gesprächen Äußerungen zu finden, welche dem bisher Gekannten widersprechen (Helfferich, 2011). Bei der Leitfadenerstellung wurde sich strikt an die allgemeinen Frageregeln und das SPSS-Prinzip nach Helfferich (2011) gehalten, womit das Grundprinzip der Offenheit gewahrt und gleichzeitig die notwendige Strukturierung im Sinne des Forschungsinteresses ermöglicht wurde. Zudem wurde jeder Leitfaden mehrmals überarbeitet. Ein maximal offener Erzählstimulus diente als Gesprächseinstieg, wie z.B.: Was ist die Geschichte, die Sie dazu gebracht hat, das Unternehmen zu gründen? Was sind Ihre Aufgaben als Director of Digital Transformation?

Zunächst wurden zwei Leitfäden erstellt, einer für jede Organisation. Diese wurden in einem zweiten Schritt an die zwei jeweiligen Interviewpartner in jeder Organisation und deren Wissen angepasst, wobei grundlegende Fragen zum Forschungsinteresse stets enthalten blieben. Jeder Leitfaden bestand aus verschiedenen Dimensionen, welche die zu stellenden Fragen thematisch ordneten. Zudem wurden bei dem ersten Leitfaden für Bernhard Prokop die Dimensionen in einer zeitlichen Reihenfolge angeordnet (Vergangenheit, Gegenwart, Zukunft), wohingegen diese beim zweiten Leitfaden für Marc Goodchild einer thematischen Logik folgten (Kooperations-zusammenhang mit Tonio, Cultural Entrepreneurship, Tonios Wert für TBS / Kompetenz-steigerung). Dieses Vorgehen zielte darauf ab, den für das Forschungsinteresse passenden Ansatz herauszufinden. Nach den ersten Interviews bei TBS in London wurde sich dann entschlossen, die Dimensionen in den kommenden Interviews bei Tonio beide nach der thematischen Logik aufzubauen (Entstehungsgeschichte & Cultural Entrepreneurship, Kooperationszusammenhang mit TBS, Tonios Wert für TBS/ Kompetenzsteigerung, strategische Zukunft).

Bernhard Prokop und Marc Goodchild von TBS wurden unter anderem folgende Fragen gestellt: Wie hat Tonio versucht, dich von dessen Nutzen zu überzeugen? Was war das für ein Pilotprojekt und wie ist es abgelaufen, gab es Probleme oder ist etwas Unerwartetes passiert? Welche Ziele hat deine Organisation und welche Rolle spielt Tonio und dessen Technologie dabei? Da aus diesen Interviews bereits interessante Themen hervorgingen, konnten die Leitfäden für das nachfolgende Interview mit Tonio teilweise angepasst werden.

Florian Novak und Wolfgang Neubert von Tonio wurden unter anderem diese Fragen gestellt: Wie sieht Tonios idealer Kunde aus und wie geht Tonio auf diesen zu und überzeugt ihn von sich und der Technologie? Wie lief die Zusammenarbeit mit Turner an dem Pilotprojekt Gumball VIP ab? Was bedeutet der Erfolg des Pilotprojektes für Tonio? Auch wenn die Leitfäden der Strukturierung des Gespräches dienen, existierte stets ein Bewusstsein dafür, dass in der konkreten Interviewsituation flexibel zwischen unterschiedlichen Themenkomplexen gesprungen werden kann.

3.4. Datenanalyse

Die Analyse der Daten bestand aus drei Schritten. Zunächst wurde eine Fallbeschreibung erstellt, welche Hintergrundinformationen zu beiden Organisationen sowie zum Kooperationsablauf und den kulturellen Managementleistungen beider enthielt (Yin, 2009).

Um im Zuge des Forschungsinteresses bestmöglich darzustellen, wie die Kooperation von Tonio und TBS mit Hilfe von „Cultural Entrepreneurship“ Leistungen entstand, ablief und welche Bedeutung interne Dynamiken bei TBS für das weitere Wachstum des Start-ups hatten, wurde in einem zweiten Schritt entschieden, dass die Analyse der Ereignisse in drei zeitlich aufeinanderfolgenden Phasen zielführend ist. Die erste Phase umfasst, wie Tonio TBS als Schlüsselkunden gewinnt. Die zweite Phase betrachtet die Durchführung des Pilotprojektes mit Tonios innovativer Technologie innerhalb von TBS. Phase drei beinhaltet Tonios Umgang mit dem erfolgreichen Pilotprojekt.

In einem dritten Schritt wurden die Daten mit Hilfe der qualitativen Analysesoftware MAXQDA analysiert, wobei sich die analytische Strategie an dem Ansatz von Mayring (2015) zur induktiven Kategorienbildung orientierte. Hierbei wurde auf einer zeitlichen Ebene zunächst ein Kodier-Ansatz gewählt, der die Daten in die drei zeitlichen Phasen einordnete, wodurch zunächst eine grobe induktive Kodierung des gesamten Datenmaterials in die drei soeben benannten Kategorien stattfand (Phase 1, Phase 2, Phase 3). Diese gaben dem Rest der Analyse einen Rahmen. Stets geleitet von der Fragestellung entstanden mehrere Codes innerhalb von „Phase 1“ (z.B. Vision, Anwendungsbeispiele, Pitch), „Phase 2“ (z.B. Pilotprojekt, beharrliche Überzeugungsarbeit, Innovation als Herausforderung) und „Phase 3“ (z.B. Mediennutzung, Fortschrittlichkeit).

Es folgte ein iterativer Zirkel zwischen diesem Kategoriensystem und den dazugehörigen Codes und der Literatur, da das Kategoriensystem weiterer Verfeinerung bedurfte und

eine weitere Analyse der Daten neue interessante Erkenntnisse versprach. Daher wurden innerhalb der Phasen eins und drei neue übergeordnete Codes (Identitätsmechanismen, Assoziative Mechanismen, Organisationale Mechanismen) kodiert, welche dem Interesse an Legitimitätsmechanismen gerecht wurden. Diesen wurden die zuvor gefundenen Codes und deren Subcodes zugeordnet, wobei teilweise auch neue Codes entstanden. Ab diesem Zeitpunkt erhielt die Analyse eine leicht deduktive Logik.

Unter den Code „Identitätsmechanismen“ in „Phase 1“ fielen so beispielsweise Subcodes wie „Analogien“ (z.B. Geschäftsmodellvergleich mit FedEX) oder „Symbolische Handlungen“ (z.B. Pitch, Anwendungsbeispiele). Zudem wurde stets streng zwischen „Narrativen“ und „Symbolischen Handlungen“ unterschiedenen. Der von Tonio gehaltene Pitch in Phase eins wurde daher sowohl unter „Narrativ“ kodiert (z.B. Phantasie wecken, Problematik aus Kundensicht) als auch unter „symbolische Handlungen“ (z.B. zwei Slides, wenige Sekunden, gemeinsame Heimat, elitäres Umfeld) (siehe Tabelle 2).

Der Code „Vision“ (z.B. Probleme des Kunden lösen, Etablierung als Standard) hingegen wurde beispielsweise gänzlich unter „Narrative“ angeordnet. Diese Art des Vorgehens ließ sich auch in Phase drei anwenden. Hier konnten „Identitätsmechanismen“ (z.B. Mediennutzung als symbolische Handlung, Reichweitensteigerung und Erfolgsgeschichte als Narrative) und „Organisationale Mechanismen“ (Erfolgreiches Pilotprojekt mit prestigevollem Kunden, Fortschrittlichkeit durch Innovation) gefunden werden.

Phase zwei hingegen, welche die Durchführung des Pilotprojektes innerhalb des Medienunternehmens umfasst, wurde durch einen weiteren iterativen Zirkel aus der Literatur überarbeitet. Mit wachsendem Verständnis der Daten wurde analysiert, welchen Ursprung die Handlungen von TBS innerhalb dieser Phase hatten. Aus diesem Anspruch ergab sich der Code „Symbiotische Überzeugungsarbeit“. Diesem wurden die Subcodes „Übereinstimmungen in Überzeugungsarbeit“ (Analogien, symbolische Handlungen, Narrative) und „Abweichungen in Überzeugungsarbeit“ (Narrative) zugeordnet. Ein Beispiel für eine Übereinstimmung im Narrativ von TBS und Tonio war die Analogie des QR-Codes, welche von beiden Akteuren bei der Erklärung der Technologie gegenüber der jeweiligen Bezugsgruppe eingesetzt wurde.

Zusammenfassend wurde zum einen durch induktive Kategorienbildung, welche im Verlauf der Analyse eine leicht deduktive Logik erhielt, zum anderen durch mehrere iterative Zirkel zwischen den Kategorien und Codes und der Literatur der Fall von Tonio und TBS im Hinblick auf die Fragestellung analysiert. Durch dieses Vorgehen wurde dem Prinzip der Offenheit nicht nur im Erhebungs-, sondern auch im Analyseverfahren Folge geleistet (Helfferich, 2011).

4. Ergebnisse

4.1. Prolog

Um im Zuge des Interesses an „Cultural Entrepreneurship“ im Innovationskontext bestmöglich zu beschreiben, wie die Kooperation von Tonio und TBS entstand, ablief und welche Bedeutung interne Dynamiken bei TBS für das weitere Wachstum des Start-ups hatten, folgt die Darstellung der gewonnenen Erkenntnisse in drei zeitlich aufeinanderfolgenden Phasen. Diese richten sich nach der fundamentalen Erkenntnis, dass nicht nur Tonio, sondern auch dessen B-to-B Schlüsselkunde TBS kulturell in der Legitimationsarbeit der innovativen Audio-Decoding-Technologie des Start-ups aktiv war. Tabelle drei soll einen Überblick darüber geben, welcher Akteur seine Überzeugungsarbeit wann an welche Bezugsgruppe richtete.

Die Analyse der Daten hat ergeben, dass Tonio für den Zugang zu seinem Schlüsselkunden TBS in Phase eins gezielte „Cultural Entrepreneurship“ Leistungen erbrachte, um seine innovative Technologie in dessen Augen zu legitimieren. Hierzu gehörten symbolische Handlungen, spezielle Narrative und der Einsatz von Analogien.

In Phase zwei, in welcher ein Pilotprojekt wichtiger Bestandteil war, wurde diese kulturelle Legitimationsarbeit innerhalb des Schlüsselkunden durch eine Allianz mit der damaligen Assistenz der Geschäftsführung, Bernhard Prokop, fortgeführt, welche interne Hindernisse bezüglich der Einführung einer Innovation innerhalb seiner reifen Organisation durch beharrliche Überzeugungsarbeit überwand. Diese Allianz umfasste auch die Übernahme von den in Phase eins angebrachten kulturellen Argumenten Tonios, was der kulturellen Kooperation der Unternehmen einen symbiotischen Charakter verleiht. Prokop jedoch überzeugte seine internen Stakeholder hauptsächlich durch den Einsatz von Narrativen.

Während Phase drei nutzte das Start-up, welches Prokop zu diesem Zeitpunkt wieder als kulturellen Akteur ablöste und erneut selbst aktiv Legitimationsarbeit leistete, den Erfolg des gemeinsam mit TBS durchgeführten Pilotprojektes, um als Konsequenz die Medien darauf anzusetzen. Zudem übernahm auch Tonio ein Argument seines Kooperationspartners TBS in die eigene Überzeugungsarbeit, um sich damit bei neuen potenziellen Kunden als seiner neuen Bezugsgruppe zu legitimieren. Dies zeigt die Symbiose im Legitimationsprozess des wachsenden Start-ups Tonio auch in Phase drei. Die kulturellen Akteure in diesem Prozess waren also in den jeweiligen Phasen zunächst Tonio, dann TBS und dann wieder Tonio. Im Folgenden werden diese drei Phasen beschrieben und die sich darin befindenden kulturellen Aktivitäten beider Organisationen mit empirischen Beispielen untermauert.

4.2. Phase 1: Tonio gewinnt TBS als Schlüsselkunden

Wie bekam das Wiener Start-up Tonio Zugang zum internationalen Medienunternehmen TBS, welcher Mittel wurde sich dafür bedient? Die Daten zeigen drei Legitimationsmechanismen; Identitätsmechanismen zur Beanspruchung einer Unternehmensidentität, organisationale Mechanismen, um Struktur und Erfolge zu vermitteln und assoziative Mechanismen zum Aufzeigen von Beziehungen oder Verbindungen. Für all diese Mechanismen gibt es in den Daten Belege, die verdeutlichen, wie Tonio bei der Gewinnung von TBS als Schlüsselkunden handelte.

Tabelle 2: Beispielhafter grober Codebaum der Phase 1

Code	Subcodes			
Phase 1	Identitäts-mechanismen	Symbolische Handlungen	Pitch Anwendungsbeispiele	zwei Slides, elitäres Umfeld, Präsentationsgrundlage, Verständnisgrundlage, interaktiv etc.
		Narrative	Pitch Anwendungsbeispiele Storytelling Vision	Phantasie wecken, Problematik aus Kundensicht, Privacy Aspekt, Etablierung als Standard, Probleme des Kunden lösen etc.
	Organisationale Mechanismen	Analogien	Geschäftsmodellvergleich mit FedEX, QR-Code für Audio	
	Assoziative Mechanismen	Struktur & Erfolge Erfahrung & Prestige	Langes Bestehen durch .com-Adresse Empfehlung durch gemeinsamen Bekannten, W. Neubert als prestigevolles Organisationsmitglied im Kundengespräch	

Tabelle 3: „Cultural Entrepreneurship“ Rollenverteilung im Kooperationsprozess

	Akteur	Bezugsgruppe
Phase 1	Tonio	TBS
Phase 2	TBS (Tonio)	Interne Stakeholder
Phase 3	Tonio (TBS)	Neue potenzielle Kunden

4.2.1. Identitätsmechanismen

Symbolische Handlung: Elevator Pitch

Eine elementare Rolle bei der Überzeugung spielte ein Elevator Pitch im September 2016. Ein Elevator Pitch ist die kürzeste Präsentationsart innerhalb des Pitchens. Innerhalb von ca. 30 Sekunden soll dabei ein informativer und überzeugender Überblick über die Organisation oder die Idee gegeben werden. Der Pitch stellte eine symbolische Handlung dar, durch welche dem Träger dieser kulturellen Leistung, in diesem Falle dem Präsidenten von Turner International, Gerhard Zeiler, erstmalig vorgestellt wurde, was Tonio überhaupt ist und wozu dessen Technologie in seinem Unternehmen fähig sein könnte. Dadurch schuf Tonio sich in den Augen seines potenziellen Schlüsselkunden eine Organisationsidentität.

> „Florian Novak war zur rechten Zeit am rechten Ort und hat den, die Nummer zwei Turner weltweit für einen klassischen Elevator-Pitch gehabt, der ein paar Sekunden gedauert hat, der im Endeffekt zwei Charts gezeigt hat und gesagt hat: Das ist, was wir uns ausgedacht haben [...]“ (Dr. Wolfgang Neubert von Tonio (Abs. 41)).

Novak wendete sich dabei bewusst an eine Person in einer sehr hohen Position innerhalb des zu akquirierenden Kunden und war auf diese wenigen Sekunden zudem „gut vorbereitet“ (Florian Novak von Tonio (Abs. 30)). Außerdem suchte Tonios CEO sich mit einem Kongress ein elitäres Umfeld aus, was auch ihn und sein Start-up als Teil dieser Elite erscheinen ließ:

> „[...] kurz vorher im Auftritt von [Gerhard Zeiler] vorm Alpbach, das ist so ein bisschen das “Davos von Österreich“, wo im Sommer ein Kongress stattfindet und da habe ich sozusagen einen Elevator-Pitch machen können und der hat sich das angehört und hat gesagt: Interessant, bitte gerne Kontakt mit meinem Assistenten, Bernhard Prokop, der kennt sich bei solchen Sachen besonders gut aus [...]“ (Abs. 30).

Der Pitch öffnete Tonio die Tür bei TBS. Auch die Tatsache, dass Zeiler wie Novak selbst aus Österreich stammt und dieser seinen Pitch daher auf der „gemeinsame[n] Sprache“ (Abs. 30) halten konnte, empfand Novak als „natürlich hilfreich“ (Abs. 30). Diese Handlungen hatten außerdem starken symbolischen Wert, da sie Gemeinsamkeiten suggerierten. Darüber hinaus gab es eine Empfehlung durch einen gemeinsamen Bekannten, über den Novak gezielt den Kontakt zu Zeiler herstellen konnte.

> „Also die Situation war die: Der CEO Florian Novak, hat meinen Chef, den Gerhard Zeiler, in Tirol bei einer internationalen Konferenz getroffen, wo Gerhard gesprochen hat. Die beiden haben einen gemeinsamen Bekannten, einen ehemaligen Politiker, der jetzt in Deutschland arbeitet

> und der hat gesagt: Da gibt es einen Typen, der ist wirklich smart, hör dir das mal an.“ (Bernhard Prokop von TBS (Abs. 8)).

Symbolische Handlung: Anwendungsbeispiel
Außerdem musste Tonio sicherstellen, dass sein Schlüsselkunde eine Vorstellung davon erhielt, wie genau seine innovative Technologie funktioniert und vor allem, wie diese sich konkret in seinem Unternehmen einbringen lassen könnte. Hierfür setzte das Start-up Anwendungsbeispiele ein und kreierte dadurch anschauliche Produkte, welche ebenso wie der Pitch die Identität Tonios in den Augen von TBS manifestierte. Die Relevanz, dass dem Kunden „zumindest die Idee“ (Florian Novak von Tonio (Abs. 20)) einer möglichen Nutzung der Technologie gegeben werden musste, wurde dem CEO und Gründer selbst erst nach einiger Zeit bewusst. Bei der Überzeugung von TBS, ungefähr zwei Jahre nach dem Markteintritt, konnte das Start-up dieses Wissen jedoch in Form von speziell dafür programmierten Anwendungsbeispielen einsetzen, wodurch es „plötzlich[e] [...] Produkte [...] gegeben“ hat (Florian Novak von Tonio (Abs. 20)).

Bernhard Prokop hat sich zunächst „nicht vorstellen können, was das [die Tonio-Technologie] ist“ (Abs. 18). Für ihn war der Besuch der Website und das darauf befindliche Beispiel in Form eines Videos aus dem Radiobereich notwendig für sein Verständnis: „Dann habe ich das Beispiel gesehen, dann habe ich es verstanden, was er [sein Vorgesetzter Gerhard Zeiler] will“ (Abs. 8). Dass Tonio ein solches Beispiel auf der eigenen Website bereitstellte, stellte eine symbolische Handlung im Sinne einer Identitätsschaffung dar. Darüber hinaus setzte Tonio diese Art der Beispiele laut Prokop auch „als Grundlage der [ersten] Präsentation“ (Abs. 12) bei TBS ein. Solche Anwendungsbeispiele förderten auch das Verständnis aller anderen, bei der ersten Präsentation bei TBS anwesenden, Personen: „es war wichtig diesen Vorführeffekt bei den anderen zu haben“ (Bernhard Prokop von Tonio (Abs. 12)). Bei der Präsentation wurde außerdem nicht bloß ein Video abgespielt, sondern die zu überzeugenden Personen wurden mit einbezogen, indem die Entrepreneure diese mit eigenen Händen ein Beispiel durchführen ließen. Wolfgang Neubert (Abs. 19) beschrieb dies als:

> „plakative[n] Demonstrator[en] mit [welchem] man halt dieses fast schon Zauberkunststück vorführt, also wenn sie halt eine Sendung sehen und das Smartphone halt diesen grünen Punkt oben hat und dann plötzlich am Smartphone das gleiche Bild erscheint, wie man es über Fernsehen sehen oder grad am Fernsehen, jetzt eine Frage entsteht, und das Smartphone sagt: Bitte tippe eins, zwei oder drei, was meinst du, was die richtige Antwort ist?“.

Durch solche Anwendungsbeispiele handelte Tonio zudem symbolisch, indem dadurch auch die Funktionalität seiner Technologie glaubwürdig unter Beweis gestellt wurde.

Narrativ: Problematik aus Kundenperspektive
Der Pitch wies jedoch nicht nur eben beschriebene symbolische, sondern auch narrative Elemente auf, mit welchen Tonio seine Identität verdeutlichte, um dadurch im Endeffekt Kunden wie TBS von sich und seinem Nutzen zu überzeugen. Dies tat Novak, indem er sich in seinem Narrativ in die Situation des Kunden hineinversetzte, um aus dessen Perspektive heraus die Problematik zu beschreiben, in welcher sich dieser aktuell befindet. Auch wenn aus den Daten nicht eindeutig hervorgeht, dass genau diese Aussagen auch bei der Akquise von TBS getroffen wurden, so ist folgende Äußerung Novaks ein Indiz dafür, dass dabei dennoch höchstwahrscheinlich sehr ähnlich vorgegangen wurde, da Novak von Broadcastern in der Fernsehindustrie sprach und auch TBS ein solcher Broadcaster ist:

> „[...] aber ich glaube sozusagen, was noch viel substanzieller ist, neben: Welchen Wert schaffen wir? Ist eben tatsächlich die Leistung, zu sagen: Wenn wir dich richtig verstehen, dann ist die Fernsehindustrie enorm, sozusagen, bedroht davon, dass die großen Player in Silicon Valley Ihnen [den Broadcastern] die Show stehlen.“ (Abs. 44).

Im nächsten Schritt des Narrativs wurde dann die Technologie als Lösung präsentiert, denn kurz nach dieser Aussage führte Novak aus, dass das Start-up für die Broadcaster in dieser Situation einen Wert schafft, indem es sie dabei unterstützt, „dass sie eben relevant am Smartphone sind“ (Abs. 44). Die Kundenperspektive wurde auch durch folgende Aussage Novaks zur Pitchsituation deutlich:

> „[...] wir sehen hier, aus Kundensicht, ja wir verlieren derzeit unsere Kunden an andere Anbieter [...] Und wir wollen einen Betrag leisten, dass diese Aufmerksamkeit wieder zurückkommt zum Broadcaster und das durchzudeklinieren, das ist sozusagen ein Schlüssel für uns.“ (Abs. 20).

Narrativ: Anwendungsbeispiel
Es liegt zudem nahe, die von Tonio eingesetzten Anwendungsbeispiele auch als Narrativ einzuordnen. Besonders deutlich wird dies in der folgenden Aussage Wolfgang Neuberts, welche das strategische Vorgehen Tonios und dessen intensive Auseinandersetzung mit der Bezugsgruppe herausstellt:

> „Das technologische Produkt als solches ist zu sperrig, ist zu technologisch, ist zu schwierig, das findet nicht den Weg in die Entscheidungsgremien der Broadcaster, sondern Sie müssen das in eine Broadcastwelt übersetzen, das heißt Sie müssen ein Produkt aufzeichnen, dass den Broadcaster abholt, das seine Pain Points bedient, und seine Sprache spricht, sein Vokabular nutzt, sein System bespielt und dann versteht er es.“ (Abs. 31).

Hier wird erkennbar, dass Anwendungsbeispiele von Tonio auch genutzt wurden, um die Bezugsgruppe dieser kulturellen Überzeugungsarbeit neugierig auf deren Technologie zu

machen. Die Notwendigkeit dessen ergab sich aus der Innovation selbst, welche „eine passive Technologie“ (Bernhard Prokop von Tonio (Abs.10)) ist und es den Kunden daher selbst obliegt, wie sie diese einsetzen. Tonio war sich dieser Herausforderung durchaus bewusst und gestaltete mehrere konkrete Anwendungsbeispiele, um die Phantasie seiner potenziellen Kunden anzuregen. Diese wurden zusätzlich durch Übersetzungsleistungen, wie Novak (vgl. Abs. 32) und Neubert (vgl. Abs. 31) es benannten, in die Welt der Bezugsgruppe übersetzt und an dessen Probleme angepasst. Dies erreichte das Start-up, indem:

> „man eben einen Broadcaster mit Beispielen konfrontiert, die er als Programmacher versteht [...] [und] nicht auf einer Metaebene über Kommunikationskanäle und bidirektionale ich weiß nicht was zu erzählen, sondern zu sagen: Stell dir vor, du hast eine, wie bei Gumball eine Kindershow und es ist immer das große Problem, dass die Kinder parallel halt am Smartphone rumtippeln und halt etwas Anderes machen und nicht zuschauen“ (Wolfgang Neubert von Tonio (Abs. 33)).

Narrativ: Innovations- und Wachstumsvision
Neben dem Pitch und den Anwendungsbeispielen zeigen die Daten, dass Tonios Narrativ aus Storytelling bestand, um dadurch eine Identität von sich für Kunden wie TBS zu schaffen. Im Zuge dessen schuf Tonio eine ausdrucksstarke Vision von seiner Technologie als „Perspektivgeber in der Not“, mit großem Potenzial, sich weltweit als akustischen Kommunikationsstandard zu etablieren. Teil dieser Vision war es, gemeinsam mit den Kunden eine zukunftsorientierte „Überlegung, was es heißt, wenn man unsere Technologie implementiert“ (Florian Novak von Tonio (Abs. 18)) anzustellen. Ergebnis dieser Überlegung war es, den Nutzern der Technologie eine Relevanz am Smartphone ihrer Kunden zu ermöglichen, was genau das war, was auch die Bezugsgruppe dringend erreichen wollte: „Und wie kann lineares Fernsehen oder unser Legacy Business mehr und mehr oder immer näher an den Konsumenten heranrücken? Und hier ist Tonio eine mögliche Antwort.“ (Bernhard Prokop von TBS (Abs.6)). Wie genau diese Relevanz am Smartphone aussehen und welchen Wert Tonio für TBS schaffen könnte, inklusive einem in der Medienbranche bisher nicht möglichen „Know-your-customer“-Effekt (Abs. 21), wurde laut Wolfgang Neubert mit Hilfe eines Business Plans und Was-wäre-wenn-Fragen gemeinsam diskutiert (vgl. Abs. 45).

Ein weiterer Teil der Vision war es, die Innovation als „völlig neuen Kommunikationskanal [zu] positionieren, der [...] aus so alten Denkmuster[n] raussteigt“ (Wolfgang Neubert von Tonio (Abs.19)). Dieses innovative Bild sah Bernhard Prokop gern, da Tonio seiner Organisation so ermöglichte, „[sich] ein bisschen schneller zu bewegen“ (Abs. 44) und daher erkannte er den Wert Tonios unter anderem darin, „ganz, ganz nah [an den Konsumenten] heranzukommen und theoretisch eine Kaufentscheidung zu forcieren“ (Abs. 14). Wie Tonio dies unter anderem tat, zeigt folgende Aussage von Wolfgang Neubert:

> „[...] da gibt es schon gewisse Bewertungsphantasien und gemeinsame Visionen, die man da aufzeigt, wo das hinführen könnte. Und das ist ja auch meist die Aufgabe eines Business Plans, dass er solche Visionen zeichnet und das einfach mal modelliert [...]“ (Abs. 45).

Narrativ: Etablierung als Standard
Einen dritten elementaren Teil der Vision beschrieb Novak wie folgt: „[...] eigentlich haben wir ja die sehr weitreichende Vision zu sagen: Tonio hat das Potenzial ein globaler Standard zu werden [...]“ (Abs. 14) und stellte sich so in den Augen von TBS als „the ubiquitous thing“ (Marc Goodchild von TBS (Abs. 49)) dar. Im Zuge dessen kommunizierte das Start-up sein Ziel auch auf der eigenen Website klar: „Tonio soll sich als Standard für die Übermittlung von Informationen über Audiosignal etablieren, das ist unser Ziel.“ (EQ# 13 Tonio, 2018). Zur Vision der Standardisierung führte Wolfgang Neubert noch weiter aus:

> „[...] wenn so ein unhörbarer Code am Mikrofon aufgenommen wird, dann weiß das Smartphone, was es damit zu tun hat und das muss die, das muss die Vision sein, dass es eine allgemein gültige Kommunikationsform ist. Neben den anderen zwei Klassikern im Nahfeldbereich, Wifi und Bluetooth.“ (Abs. 65).

Narrativ: Privacy Aspekt
Außerdem kommunizierte Novak laut Prokop den Privacy Aspekt seiner Technologie klar: „Florian hat großen Wert darauf gelegt und hat gesagt: Leute, der große Unterschied zwischen Shazam und dem was wir machen ist, Shazam hört, nimmt die Wave auf“ (Abs. 48). Auch im Newsletter des Start-ups, welcher von Tonios Fortschritten in Silicon Valley berichtete, wurde dieser Aspekt als Vorteil der Technologie hervorgehoben: „Our goal is to create a global standard for sharing data via audio, placing convenience and privacy at the helm.“ (EQ# 21 Tonio, 2018).

Narrativ: Tonio als magische Technologie
Tonio ließ auch durch die Beschreibung seiner Technologie als magisch ein Bild von sich und der Innovation in den Köpfen der Bezugsgruppe entstehen. Diese befindet sich direkt auf der Homepage von dessen Website: „Demo-Quiz - Du willst die Magie von Tonio gleich erleben? Lade Dir die Tonio-App hier herunter und starte das Beispiel.“ (EQ# 13 Tonio, 2018). Auch in seinem Newsletter betonte das Start-up die „Magie seiner Technologie“ im ersten Satz: „The team at Tonio has been busy supporting our partners as they carve new niches into the media landscape and showing off the magic of tone with information.“ (EQ# 21 Tonio, 2018).

Analogien: QR-Code und FedEX
Die Daten legen zudem nahe, dass Tonio TBS seine Identität neben symbolischen Handlungen und Narrativen auch durch den Einsatz von Analogien beschrieb. Das Geschäftsmodell

seiner Organisation erklärte Novak so durch den Vergleich mit dem Logistikunternehmen FedEX: „Wir nehmen das was du schicken willst, das verpackst du, wir stellen das zu und du zahlst uns für die Zustellung.“ (Abs. 20). Eine zweite von Novak eingesetzte Analogie in der Überzeugungsarbeit bestand darin, die Technologie als „QR-Code für Audio“ (Abs. 10) zu beschreiben, um damit „das Verständnis, dass man ein bestehendes Gerät funktionell erweitern kann“ (Abs. 12), anzuregen.

4.2.2. Organisationale Mechanismen

Struktur & Erfolge

Identitätsstiftende Mechanismen machten einen Großteil der Überzeugungsarbeit Tonios in Phase eins aus. Für die Legitimationsmechanismen zwei und drei ergeben die Daten weniger ausführliche Hinweise. Dennoch, ein zweiter Legitimationsmechanismus, mit welchem Tonios Überzeugungsarbeit in der Gewinnungsphase von TBS beschrieben werden kann, ist ein organisationaler. Mit dessen Hilfe vermittelte Tonio potenziellen Schlüsselkunden Struktur und Erfolge, um sich selbst als erfahren darzustellen. Wie anderen Start-ups mangelte es Tonio noch an Leistungsnachweisen, was dem Gründer Florian Novak jedoch bewusst war und er so bereits in der Gründungsphase Maßnahmen ergriff, um ein vertrauenswürdiges Bild von Tonio am Markt zu schaffen. Dies tat er, indem er ein langes Bestehen seiner Organisation vermittelte: „marketingmäßig habe ich gewusst, ich brauche eine .com-Adresse, es muss sozusagen so daherkommen, als würde es das [Tonio] schon immer geben.“ (Abs. 2).

4.2.3. Assoziative Mechanismen

Beziehungen zur Elite

Der dritte und letzte Legitimationsmechanismus beruhte auf Assoziation, womit Tonio TBS vermittelte, welche vertrauenswürdigen Beziehungen oder Verbindungen das Start-up hatte. Hierzu zählt die eingangs beschriebene Empfehlung an Gerhard Zeiler, den Präsidenten von Turner International, durch einen Politiker, wodurch Tonio gute Beziehungen zu gesellschaftlich anerkannten und einflussreichen Persönlichkeiten demonstrierte: „Da gibt es einen Typen, der ist wirklich smart, hör dir das mal an.“ (Bernhard Prokop von TBS (Abs. 8)), zitierte Prokop die Worte des gemeinsamen Bekannten an seinen Vorgesetzten.

Prestigevolle Organisationsmitglieder

Außerdem befand sich mit Wolfgang Neubert auch innerhalb von Tonio ein durchaus prestigevolles Organisationsmitglied, welches Novak „bei wichtigen Gesprächen mit unserem Kunden Turner von Anfang an eingebunden [hat]“ (Abs. 8) und dem Medienunternehmen dadurch strategische und fachliche Kompetenz, eine gewisse Erfolgsgarantie sowie sehr gute Vernetzung und Einfluss in der österreichischen Wirtschaft versprach. Darüber hinaus kommen Novak, Neubert, Zeiler, der Präsident von Turner International und auch Prokop alle aus Österreich, was der Kooperation zumindest nicht im Weg gestanden haben sollte.

Zusammenfassend hat Tonio in der Rolle eines kulturellen Akteurs vielfältige kulturelle Überzeugungsleistungen, hauptsächlich bestehend aus Identitätsmechanismen und gerichtet an seinen Schlüsselkunden TBS, als Bezugsgruppe dieser Leistungen, erbracht. Wie genau diese Handlungen in die finale Entscheidung von TBS für die Kooperation mit dem Start-up eingeflossen sind, kann an dieser Stelle nicht klar definiert werden, jedoch zeigt die Analyse eindeutig, dass Tonio im Sinne von „Cultural Entrepreneurship“ bewusst und aktiv gehandelt hat. Die erste Phase des Legitimationsprozesses wurde somit dargestellt.

In der zweiten Phase, welche im nächsten Teil thematisiert wird, fand eine Umverteilung der Rollen statt. Tonio und dessen Technologie waren zu diesem Zeitpunkt bereits erfolgreich bei TBS legitimiert und die Kooperation war beschlossen – ab diesem Zeitpunkt wurde Bernhard Prokop kulturell aktiv und überzeugte den Rest seiner Organisation von Tonios Technologie, ebenso, wie Tonio ihn und seine Organisation zuvor überzeugte. Die Rolle der Träger von Prokops kultureller Überzeugungsarbeit nahmen in Phase zwei daher die Mitarbeiter von TBS als interne Stakeholder ein.

4.3. Phase 2: Pilotprojekt bei TBS – Interne Allianzpartner schaffen für Tonio Akzeptanz

Wie unterstützte TBS Tonio also dabei, Akzeptanz für die innovative Technologie zu schaffen und ein erstes Pilotprojekt bei sich mit dieser durchzuführen? Hierzu ergeben die Daten, dass insbesondere Bernhard Prokop mobilisiert wurde und intern beharrlich Überzeugungsarbeit für Tonio leistete, wobei er sich der Herausforderung einer Innovationseinführung in einer reifen Organisation wie TBS stellte. Ein ausschlaggebendes Mittel dafür waren PowerPoint Präsentationen. Da Prokop Teile von Tonios Legitimationsmechanismen aus der ersten Gewinnungsphase (Symbolische Handlungen, Narrative und Analogien zur Identitätsschaffung) in die eigene, an interne Stakeholder gerichtete Argumentation übernahm und es zudem, als damaliger Assistent der Geschäftsführung, als mit einem potenziellen Aufstieg verbundene Chance begriff, sich durch diese kulturelle Überzeugungsarbeit als Innovationstreiber innerhalb von TBS zu positionieren, stellt diese Allianzsituation nicht nur eine Partnerschaft, sondern eine Symbiose dar. Zunächst soll der weitere Verlauf der Ereignisse rund um die Kooperation der beiden Organisationen beschrieben werden, um danach als Ergebnis der Analyse zu zeigen, wie Prokop, hauptsächlich durch den Einsatz von Narrativen, seine Überzeugungsarbeit gestaltete.

4.3.1. Tonios innovative Technologie als Herausforderung für TBS

Da für die Nutzung der Tonio-Technologie von TBS selbst passende Inhalte geschaffen werden mussten, was ein Umdenken in den Köpfen der Mitarbeiter erforderte, bestanden innerhalb von TBS von Anfang an Zweifel. Dies beschrieb Bernhard Prokop wie folgt:

> „Also man hat sehr viel kreativen Freiraum, aber der kreative Freiraum ist sehr kritisch, den muss man auch nutzen und da sind große multinationale Unternehmen jetzt nicht gerade die besten, die sich da auf neue Gebiete vorwagen.“ (Abs. 24).

Zudem berichtete er, dass die Mitglieder innerhalb großer Organisationen wie TBS am liebsten „das [tun], was sie am besten können, nämlich das was sie heute tun und nicht was dann möglicherweise auf uns zukommen würde.“ (Abs. 19). Dies zeigte sich auch durch Prokops Beschreibung der ersten Reaktionen auf die Innovation: „So was ist das? Was tut das? Wie schaut das aus? Und da muss man sagen, viele Leute haben gleich gesagt: Uh. Das ist ziemlich tough das umzusetzen [...]“ (Abs. 10).

Dennoch, die Mission von TBS war es, das Fernsehen neu zu erfinden (EQ# 18 TBS, 2019) und nachdem Tonios Technologie im Frühjahr 2017 durch umfangreiche Tests seine Funktionalität bewies, entschloss sich das Medienunternehmen, ein auf dieser Technologie basierendes Pilotprojekt durchzuführen: „In order to see the technology's impact Turner decided create a Pilot which involves the technology.“ (EQ# 6 TBS, 2017, S. 2). Bei dem Pilotprojekt handelte es sich um die „Gumball Superfan Experience“, eine „nine day TV marathon second-screen app experience [...]“ (EQ# 9 TBS, 2018, S. 2). Diese wurde von April bis Mitte Oktober 2017 unter der operativen Führung von Marc Goodchild entwickelt. Die Kinderserie „Die fantastische Welt von Gumball“ lieferte dafür die Inhalte, welche aus den Abenteuern des blauen Katers Gumball bestehen, der bei seinen alltäglichen Streichen von einem Goldfisch unterstützt wird. Die Idee für das Pilotprojekt bestand darin, den Zuschauern eine Möglichkeit der ausstrahlungssynchronen Rückmeldung durch die Beantwortung von Quizfragen zum Inhalt innerhalb der Gumball-App zu geben. Gewonnene Punkte konnten im Anschluss auf einer Gumball-Mikrowebsite des Senders in Fanartikel eingelöst werden (EQ# 9 TBS, 2018). Hierfür wurde den Zuschauern die klare Nachricht „Tune in to Gumball episodes every day at five o'clock“ (Marc Goodchild von TBS (Abs. 28)) übermittelt.

Die Reaktionen der Mitarbeiter auf dieses neuartige Wiederholungskonzept waren verhalten, dennoch machte Goodchild seinem Team klar, dass ein starkes Umdenken zwingend erforderlich war:

> „We were asking them to throw that thinking in the bin and go: No, you are going to have to do it the same time every day, because we need a clear message [...]“ (Abs. 32).

Um das Pilotprojekt durchzuführen, war darüber hinaus eine abteilungsübergreifende Zusammenarbeit notwendig, obwohl es in großen Organisationen in der Medienbranche „sehr viele Silos [gibt], es ist sehr “ich mach das und du machst das“ und da gibt es keine notwendige Kooperation“ (Bernhard Prokop von Tonio (Abs. 22)). Die Innovation in seiner Organisation zu etablieren, war für Prokop also kein Leichtes. Dies zeigt auch die folgende Aussage, in welcher die Trägheit der Mitarbeiter während der Implementierung deutlich wird:

> „Ja, so haben wir das [Pilotprojekt] angefangen und es war auch in der Organisation, es war nicht, dass jeder gleich gedacht hat: Super das machen wir!“ (Abs. 22).

Kurz danach beschrieb Prokop, wie Versuche, die Mitarbeiter von der Innovation zu überzeugen, teilweise sogar scheiterten:

> „Schaut her, das könnten wir machen, das wäre was Neues. Würde euch das nicht interessieren? Die konnten sich das dann alles nicht vorstellen. Dann hatten verschiedene Abteilungen wieder keine Zuständigkeit und sagten: Mir ist es völlig egal, ob es das tut oder nicht. Ich mache nur das.“ (Abs. 22).

Parallel zur Implementierung der Serieninhalte in die App mit Hilfe von Tonios Software Development Kit (SDK), wurden Werbekampagnen durchgeführt, um das Pilotprojekt anschließend vom 21.10.2017 bis zum 29.10.2017 in Großbritannien live zu schalten (EQ# 6 TBS, 2017). Vom großem Erfolg dieses ersten Tests war Prokop anschließend selbst „ein bisschen überrumpelt“ (Abs. 22), da „Gumball VIP“ sowohl Auswirkungen auf die sonst „sehr unpräzise[n]“ (Abs. 22) Ratings, als auch auf das „Engagement“ der Zuschauer hatte (EQ# 10 TBS, 2018). So ergaben Auswertungen eine eindeutige Korrelation zwischen dem „Gumball VIP app engagement“ und der „on-air activity“ sowie ein hohes „average enagagement“ von neun Minuten während einer Folge, welche insgesamt elf Minuten dauerte (EQ# 10 TBS, 2018).

Sowohl der interne Widerstand gegen die Arbeit mit Tonios innovativer Technologie als auch das grundlegende Desinteresse an dessen Integration blieb bestehen und weder das Pilotprojekt noch Prokop erhielt internen Zuspruch:

> „Ja, es hat mich dann schon überrascht, dass es wirklich in Wahrheit niemanden gibt, den es wirklich interessiert hat, [...] es ist nur quasi durchgeführt worden, weil ich die ganze Zeit nervig war [...] Es hat keiner die Notwendigkeit gesehen.“ (Abs. 26).

So handelte Prokop nach der Feststellung des Erfolges folgendermaßen: „[...] für mich hat es bedeutet, dass ich weiterhin ein pain in the ass sein kann und [...] auf Basis davon habe ich meine amerikanischen Kollegen überzeugt, dass das eigentlich etwas ist, dass sie ausprobieren sollten.“ (Abs. 32; EQ# 22 der Standard, 2017). Dies verdeutlicht, dass Prokop, anstatt sich entmutigen zu lassen, diesen ersten Erfolg nutzte, um die Arbeit mit Tonios Technologie auszubauen und diese so stetig weiter zu legitimieren:

> „Du brauchst in einer Organisation das Erste. Alle rennen und tun, das Erste, ja jetzt haben

> wir das Erste. Und jetzt könnte ich sagen: Super, lässig gemacht, danke, Wiederschauen! Da sage ich: Nein, jetzt machen wir noch ein Zweites und vom Zweiten machen wir jetzt auf einmal zwar was Ähnliches, aber ein größeres Volumen in zehn, mit zehn Programmen. Sind zehn Programme global? Ist das ansatzweise 24 Stunden, live? Nein, es sind zehn. Was ist der nächste Schritt? Vielleicht eine Stunde, oder ein Nachmittag und was kommt dann, und was kommt dann, und was kommt dann?“ (Abs. 32).

So folgte die Ausweitung der „Gumball Superfan Experience“ auf zehn weitere Märkte (EQ# 23 der Standard, 2018). Marc Goodchild beschrieb die Reaktionen der Mitarbeiter in den verschiedenen Ländern folgendermaßen: „[...] every one of those markets also have their own politics, quite a lot of people are naturally conservative. You know, it is sort of: Uh, this looks scary.“ (Abs. 32). In Italien, auf den Philippinen und Singapur ging „Gumball VIP“ im November 2018 auf Sendung, im Frühjahr 2019 folgen unter anderem Australien, Rumänien und Frankreich (EQ# 12 TBS, 2018).

Nachdem die Ereignisse rund um das Pilotprojekt sowie die damit einhergegangenen Schwierigkeiten für die internen Allianzpartner Prokop und Goodchild wiedergegeben und deren maßgeblicher Einsatz für die Arbeit mit der Technologie verdeutlicht wurde, werden nun die Analyseergebnisse dieser zweiten Phase vorgestellt. Diese zeigen die geleistete Überzeugungsarbeit, mit welcher hauptsächlich Prokop die Technologie in den Augen interner Stakeholder legitimierte und so die Rolle des kulturell aktiven Teil-Entrepreneurs im symbiotischen Kooperationsprozess des Start-ups Tonio und dessen B-to-B Schlüsselkunden TBS übernahm. Hierbei hat er stets eigene Wege gefunden, ist an „Barrieren vorbeigekommen [...] und [hat] nicht locker [ge]lassen“ (Bernhard Prokop von TBS (Abs. 10)), sah von Anfang an das Potenzial des „Second Screens“ und kommunizierte dies unter anderem in PowerPoint Präsentationen. Wie bereits beschrieben, bestand Tonios Legitimitätsarbeit hauptsächlich aus Identitätsmechanismen (Symbolische Handlungen, Analogien, Narrative). Prokops Überzeugungsarbeit bestand ausschließlich aus solchen Identitätsmechanismen und er nutzte hauptsächlich Narrative, um ein Bild von der innovativen Organisation Tonio und der Technologie in den Augen seiner internen Stakeholder bei TBS entstehen zu lassen und diese so trotz fehlender Motivation und teilweise sogar geleistetem Widerstand von der Innovation zu überzeugen.

4.3.2. Identitätsmechanismen

symbolische Handlung: Anwendungsbeispiele

Prokop handelte in seiner Legitimationsarbeit ebenso wie Tonio zuvor symbolisch. Dies tat er, indem er Anwendungsbeispiele in seiner Überzeugungsarbeit einsetzte, um seiner Bezugsgruppe Tonios Funktionsweise zu erklären und die Möglichkeiten der Technologie aufzuzeigen (EQ# 5 TBS, 2016; EQ# 7 TBS, 2018; EQ# 8 TBS, 2018). Wie wichtig solche Beispiele für das Verständnis der Bezugsgruppe sind, konnte Prokop schon bei der ersten Präsentation von Tonio beobachten:

> „[...] sie sind gekommen [...] und hatten eine Präsentation vorbereitet. Es gab darin keine Fülle von Anwendungsbeispielen aber eines oder zwei und die waren auch die Grundlage der Präsentation. Ich habe es schon verstanden, bevor sie es mir gezeigt haben, aber es war wichtig diesen Vorführeffekt bei den anderen zu haben.“ (Abs. 12).

symbolische Handlung: Adaption an TBS

Diese Anwendungsbeispiele passte er zudem exakt an TBS an. So zeigen die Daten, dass Prokop ein Anwendungsbeispiel brachte, in welchem CNN, einer der Sender von TBS, als Broadcaster die Technologie einsetzte (EQ# 7 TBS, 2018). Ein anderes Beispiel zeigte die TV-Sendung „Next Top Model“ (EQ# 7 TBS, 2018). Auch in einer der ersten internen Präsentationen über Tonio, welche Ende 2016 von Gerhard Zeiler gehalten wurde, befanden sich diverse, auf die Branche zugeschnittene Anwendungsbeispiele: „Range of application within our media world“ (EQ# 5 TBS, 2016, S. 16). Tonios anfängliches Anwendungsbeispiel hingegen, mit welchem Bernhard Prokop erstmalig von der Technologie erfuhr, stammte wie bereits beschrieben aus dem Radiobereich und war daher weniger gut auf die Bezugsgruppe abgestimmt (siehe Phase 1: 1.2).

Narrativ: Adaption an TBS

Zudem passte Prokop auch sein Narrativ exakt an die Fernsehbranche beziehungsweise konkret an TBS an. Dies erreichte er auf verschiedenen Wegen. Zum einen schuf er für die internen Stakeholder Produkte und sogar ganze Visionen. Eine solche zukünftige Verwendung von Tonio beschrieb er in einer PowerPoint Präsentation, welche weitere Einsatzmöglichkeiten der Technologie vorstellte: „After the successful launch and implementation of the „Gumball VIP“ Pilot [...]“ (EQ# 6 TBS, 2017, S. 9), schlug er vor, neben der weniger aufwendigen und ressourcenschonenderen Duplizierung des Pilotprojektes in weitere Märkte (wozu sich letztendlich auch entschieden wurde), einen „Cartoon Network Loyalty Club“ zu entwickeln:

> „The vision is to develop and create an interactive CN fan destination, which connects linear TV (CN channel) with a digital CN destination (club) in which the kids who watch TV (linear or digital) and use the CN club app get rewarded (reward points).“ (EQ# 6 TBS, 2017, S. 13).

Dies demonstriert, dass Prokop die weitere Zusammenarbeit mit Tonio aktiv intern vorantrieb. Zudem sprach er auch klare Handlungsempfehlungen für den „CN Loyalty Club“ als „360°brand experience“ aus, da diese einen „New innovative approach of content and audience interaction“ darstellte, womit er sogar die Zahlung weiterer Lizenzgebühren an Tonio rechtfertigte (EQ# 6 TBS, 2017, S. 17).

Narrativ: Erfolg des Pilotprojektes
Prokops vielschichtiges Narrativ zeichnete sich auch durch den Fokus auf den Erfolg des Pilotprojektes „Gumball VIP“ aus. Eine solche Erfolgsbilanz konnte Tonio in Phase eins nicht vorweisen. Welchen Wert Prokop darauf legte, Tonios Geschichte innerhalb von TBS als Erfolgsgeschichte zu erzählen, geht eindeutig aus einer seiner PowerPoint Präsentationen hervor, in welcher er die Performance des Pilotprojektes in Form von diversen Kennzahlen zusammenfasste und schließlich folgenden Satz betonte: „Gumball VIP was a success“ (EQ# 10 TBS, 2018, S. 2). In dieser Präsentation verglich er „Gumball VIP“ zudem mit anderen Formaten des Medienunternehmens, wobei diese in den meisten Fällen schlechter abschnitten als das auf der Innovation basierende Pilotprojekt. Zudem sagte Prokop, dass dessen Erfolg ihm als Basis diente, um auch die Kollegen in den USA davon zu überzeugen, „dass das eigentlich etwas ist, dass sie ausprobieren sollten“ (Abs. 32) und dass die Akzeptanz für die Technologie seiner Meinung nach als entscheidender Faktor durch den Erfolg gestiegen ist (vgl. Abs. 32). Prokop als Allianzpartner stoppte seine Überzeugungsarbeit nach dem ersten Erfolg also nicht, sondern weitete sie in London nach Cartoon Network auf andere Sender wie CNN oder sogar weitere Länder aus.

Narrativ: Innovations- und Wachstumsvision
Ein weiterer Teil des Narrativs, mit welchem Prokop Überzeugungsarbeit für Tonio leistete, bestand darin, dass er die innovative Technologie als Chance oder auch Mode im Kontext der Digitalisierung darstellte. Dies wird insbesondere durch folgende Aussagen aus PowerPoint Präsentationen Prokops deutlich: „Audio is the new touch according to science articles and recent product development (Alexa, Siri, Google Home)“ (EQ# 7 TBS, 2018, S. 16) sowie „Turner can set a new trend and establish a new way of thinking and interacting with (TV) content and their audience (!)“ (EQ# 6 TBS, 2017, S. 17). Zudem beinhaltete die erste Präsentation mehrere „innovative use cases“ (EQ# 5 TBS, 2016, S. 19), wie zum Beispiel einen „Smart Mirror“ in welchem die Tonio-Technologie eingesetzt werden könnte. Hiermit traf Prokop den wunden Punkt des Medienunternehmens, welches in seinem „legacy business sehr veraltet und klassisch und ja nicht besonders innovativ“ (Bernhard Prokop von TBS (Abs. 22)) war und sich bezüglich Innovation zu dem Zeitpunkt „gerade mal in der Kriechposition“ (Bernhard von TBS (Abs. 44)) befand. Darüber hinaus war die Beschreibung einer möglichen Zukunft des Medienunternehmens mit der Tonio Technologie beziehungsweise die Lenkung der Vorstellungen der Bezugsgruppe in eine Richtung, in welcher Tonio relevant oder von Nutzen sein könnte, Teil des Narratives von Prokop. Ähnliche Elemente wies auch die Vision Tonios auf, welche das Start-up in der vorherigen Phase an TBS kommunizierte (siehe Phase 1: 1.5).

Genauso war es ein wichtiger Teil von Prokops Präsentationen, Möglichkeiten mit der Technologie aufzuzeigen (EQ# 5 TBS, 2016; EQ# 6 TBS, 2017; EQ# 7 TBS, 2018), was sich auch durch Aussagen wie „Imagine future“ (EQ# 7 TBS, 2018, S. 11) oder „Next steps: re-imagine TV together with Tonio“ (EQ# 5 TBS, 2016, S. 21), zeigt.

Narrativ: Technologie als Hilfsmittel für Fokus auf Kerngeschäft
Darüber hinaus kommunizierte Prokop den Nutzen der Technologie, indem er diese als Hilfsmittel für TBS und die gesamte Medienbranche zur Rückbesinnung auf deren eigentliches Kerngeschäft beschrieb: „Let’s agree on technology (we as Turner should participate though) and compete on content“ (EQ# 7 TBS, 2018, S. 16). An dieser Stelle zeigt die Datenlage nicht eindeutig, ob dieser Satz wirklich von Prokop stammt, doch da Novak laut eigener Aussage nicht der Urheber war, liegt es nahe, dass Prokop derjenige gewesen ist, was wiederum eine weitere eigenständige kulturelle Überzeugungsleistung durch den Schlüsselkunden selbst bedeutet. Anders als das Technologie Start-up, welches in seinem eher auf Umbruch basierenden Narrativ die Relevanz am Smartphone oder die Etablierung der Technologie als Standard hervorhob, motivierte und überzeugte Marc Goodchild seine Mitarbeiter darüber hinaus mit einer eigenen, zur Fernsehbranche passenden, Argumentation:

> „So my internal sell has been about: This is about driving back to linear TV, it is about helping people fall in love with the first [...] screen and the reason that is important, which is, how I sell it in internally, is that we still command a large share of our advertising revenues of TV.“ (Abs. 36).

Narrativ: Tonio als magische Technologie
Prokop versuchte seine Bezugsgruppen außerdem von Tonio zu überzeugen, indem er die Technologie als magisch umschrieb. Dies wird in einer seiner Präsentationen deutlich, in welcher er ein Beispielvideo verwendete und dafür die Überschrift „JUST MAGIC“ (EQ# 7 TBS, 2018, S. 3) wählte. Dasselbe tat Marc Goodchild während der Implementierung der Technologie: „I keep on talking to them about: Tonio is magic.“ (Abs. 36). Wie zuvor erwähnt, beschrieb auch Tonios Narrativ dessen Innovation als magisch, um sich bei der Bezugsgruppe TBS zu legitimieren (siehe Phase 1: 1.8).

Narrativ: Privacy Aspekt
Ein weiterer Teil der kulturellen Überzeugungsarbeit von Bernhard Prokop als Allianzpartner Tonios bestand aus der Betonung des Schutzes von Tonios Nutzern und von deren Privatsphäre. Dieselbe Argumentation setzte auch Tonio in Phase eins (1.7) ein, um TBS von sich zu überzeugen. Dass auch Prokop wenig später den Sicherheitsfaktor der Technologie in seiner Legitimationsarbeit einsetzte, zeigten seine Präsentationen (EQ# 5 TBS, 2016; EQ# 7 TBS, 2018; EQ# 8 TBS, 2018) sowie seine Äußerung im Interview:

> „Und was für mich schon im letzten Jahr interessant war oder einen zusätzlichen Nutzen gegeben hat, war die Konversation rund um das Thema Sicherheit. Es ist eine sehr sichere, also eine sehr auf Privacy fokussierte Technologie und ich glaube, dass das ein großer Vorteil gegenüber Alexa und wie sie alle heißen [...]“ (Abs. 46).

Narrativ: Etablierung als Standard
Ein letztes Narrativ, welches Prokop in der Überzeugungsarbeit einsetzte, stellte Tonios Ziel dar, sich als weltweiten Standard zu etablieren. Wie Tonio dies tat, wurde bereits unter Phase eins (1.6) ausführlich beschrieben und die Daten zeigen, dass auch Bernhard Prokop Tonios Vision der Standardisierung in seine Kommunikation integrierte. In der ersten Präsentation, in welcher Tonio TBS vorgestellt wurde, stellte Prokop die Frage, wieso das Medienunternehmen Tonio kennen sollte. Diese beantwortete er selbst, indem er Tonios Vision, sich als „global technology standard" (EQ# 5 TBS, 2016) zu etablieren, vorstellte. Hierbei nutzte er die Symbole von den bereits standardisierten Technologien Wifi und Bluetooth und platzierte Tonios Logo auf einer Ebene neben diesen. Auch in anderen Präsentationen beschrieb Prokop Tonios Ziel eindeutig: „Tonio's vision is a global standard around specific „watermarking technology" (Audio QR-Code) [...] The integration into the iOS and Android will follow soon [...] (EQ# 8 TBS, 2018, S. 5).

Abgesehen von Prokops Narrativ, verdeutlichen die Daten, dass dieser Tonio auch auf Grund eines eigenen Interesses an dessen weltweiter Etablierung unterstützte:

> „[...] wenn sich das [Tonios Technologie] als Standard etabliert, dann hat das gute Auswirkungen auf mich und auf unser Unternehmen und wenn wir einen „Minority Stake" haben von 25%, 30%, dann ist der Wert auf einmal nicht X sondern ein paar Milliarden und dann habe ich einen „Return on Investment" generiert fürs Unternehmen vom fünf Millionenfachsten. Und das ist nicht gerade, das ist jetzt kein Karrierehindernis, sagen wir mal so." (Abs. 56).

Im Zuge dessen verweisen die Daten teilweise auch darauf, dass Prokop für Tonio den Kontakt zum Technologiekonzern AT&T herstellte, welcher im Juni 2018 WarnerMedia, die Mutter von TBS, übernahm. Beispielsweise befindet sich in den unternehmensinternen Daten ein Dokument mit der Überschrift „How AT&T can benefit from employing Tonio." (EQ# 11 TBS, 2018), in welchem Florian Novak dem Konzern offiziell seine Kontaktdaten hinterlässt und mögliche Anwendungsfälle beschreibt, durch welche AT&T zusammen mit TimeWarner diverse Vorteile des „upcoming new acoustic communication standard" genießen könnte. Prokops Bereitschaft, sich für eine Standardisierung der Innovation einzusetzen, wird zudem durch folgende Aussage verdeutlicht:

> „Wenn Tonio mit Warner Media und AT&T Erfolge erzielt, dann könnte man die Jungs schon dazu zwingen, dass sie das implementieren müssen. Das ist jetzt nicht ganz so unrealistisch, aber es lastet leider sehr viel auf meinen Schultern, ob das zustande kommt oder nicht." (Abs. 54).

Dementsprechend äußerte Prokop ein „vitales Interesse" (Abs. 64) daran, die Zusammenarbeit mit Tonio nach „Gumball VIP" auch in Zukunft weiter auszubauen. Doch auch die Medienbranche hätte laut Prokop letzten Endes großes Interesse daran, „Technologie zu standardisieren" (Abs. 14), um damit:

> „endlich in den Bereich zurückzukehren, wo wir eigentlich Konkurrenten und Konkurrentinnen sind, nämlich im Contentbereich. Wir sind keine Technologieunternehmen, das sind wir nicht. Aber wir sind derzeit gezwungen auf Technologieebene zu konkurrieren, weil keiner Antworten hat auf die revolutionierte Distribution von Content, die Netflix und Amazon eingeführt haben." (Abs. 14).

Doch die Aussicht auf Standardisierung war für TBS nicht zwingend nötig, um Tonios Technologie bei sich einzusetzen, da diese zunächst auch die Konkurrenzfähigkeit des Medienunternehmens verbesserte: „hier ist Turner und die meisten großen Medienunternehmen sicher noch nicht dort wo die großen Direct-to-Consumer Anbieter so wie Netflix oder Amazon heute sind." (Bernhard Prokop von TBS (Abs. 4)). Auch dessen Kollege Marc Goodchild betonte:

> „The new players are over here, completely separate, we are tying to make sure we are prepared. So one the one hand, dominate in what we are doing, on the other hand is, make sure we are in the new space and able to compete before they get too big." (Abs. 8).

TBS war sich also bewusst, dass Tonio dessen digitale Fähigkeiten steigern würde und förderte wahrscheinlich auch deswegen Tonios Technologie – unternehmensintern und branchenüber-greifend. Damit zeigt sowohl die zuvor herausgestellte Übereinstimmung diverser „Cultural Entrepreneurship" Leistungen von Tonio und TBS (Anwendungsbeispiele, QR-Code Analogie, Innovations- und Wachstumsvision, Tonio als magische Technologie, Privacy Aspekt, Etablierung als Standard) als auch die übergeordnete Hoffnung des Medienunternehmens und Prokops, darauf, von dem Erfolg oder einer möglichen Standardisierung der Technologie zu profitieren, dass die Handlungen in der zweiten Phase über eine reine Unterstützungsarbeit hinausgingen. Vielmehr war es eine Symbiose zweier Akteure, welche sich gemeinsam für die Legitimation einer innovativen Technologie im weitesten Sinne einsetzten.

Analogien: QR-Code
Wie bereits unter Phase eins beschrieben (siehe 1.9), nutzte Novak Analogien wie die des QR-Codes, um Kunden wie TBS das Geschäftsmodell seines Start-ups und die Funktionsweise seiner Technologie zu erklären: „Grundsätzlich formuliere ich es ja sehr klar, dass wir sagen: Wir schicken Daten über Audio. Die Kernanwendung ist der QR-Code für Audio [...]" (Florian Novak von Tonio (Abs. 24)). Auch Prokop nutzte diese Art des Erklärens und verwendete die QR-Code Analogie, um seinen Stakeholdern innerhalb von TBS auf die gleiche Art zu erklären, was es mit der innovativen Technologie auf sich hatte. Folgendes Zitat Prokops zeigt, dass dessen Überzeugungsarbeit durch die von Tonio inspiriert war:

> „[...] innerhalb des Unternehmens sage ich immer, mit der Technologie kann man lineares Fernsehen digitalisieren. Dann schauen immer alle ganz groß: Was könnte das heißen? Aber eigentlich müsste man es beschreiben mit: Es funktioniert ähnlich wie ein unhörbarer QR-Code, das ist das Beispiel was immer der Florian nimmt und ich kann diesem Argument immer mehr etwas abgewinnen, weil viele verstehen wie der QR-Code funktioniert [...]“ (Abs. 72).

Des Weiteren verwendete Prokop die Analogie auch innerhalb seiner PowerPoint Präsentationen (EQ# 7 TBS, 2018; EQ# 8 TBS, 2018), welche er in diesem Sinne als Instrument nutzte, um Wissen über Tonio zu produzieren und zu verbreiten.

4.4. Phase 3: Konsequenz des erfolgreichen Pilotprojektes

Wie ging Tonio mit dem Erfolg seiner Technologie bei TBS um? Nachdem in Phase zwei Tonios Schlüsselkunde die Rolle des Überzeugungsarbeit leistenden Akteurs übernahm und dabei maßgeblich zum Erfolg „Gumball VIPs“ beitrug, wurde in Phase drei, der letzten Phase im Prozess, erneut Tonio zum aktiveren Part, der sich für die Legitimation seiner innovativen Audio-Decoding-Technologie einsetzte. Die Bezugsgruppe war in dieser Phase jedoch nicht mehr ein konkreter Kunde wie TBS, sondern die Gesamtheit neuer potenzieller Kunden. Die Daten zeigen erneut Identitätsmechanismen und auch organisationale Mechanismen, durch welche Tonio versuchte, sich bei neuen Kunden zu legitimieren und so als Start-up weiter zu wachsen.

4.4.1. Identitätsmechanismen

Symbolische Handlung: Mediennutzung

Ende 2017 machten TBS und Tonio die Zusammenarbeit durch eine Pressemitteilung des Medienunternehmens öffentlich (siehe nachfolgend Punkt 2), was eine symbolische Handlung und damit einen Identitätsmechanismus darstellte. Doch die Daten machen auch deutlich, dass Tonio die Medien darüber hinaus aktiv auf diese Mitteilung aufmerksam machte, um durch mediale Präsenz weiteres Legitimitätskapital aus der Kooperation mit seinem prestigeträchtigen Schlüsselkunden zu schlagen:

> „Also der unmittelbare Anlass war eine Pressesendung von Turner. Natürlich gehört auch zur PR auf unserer Seite, dass wir solche Dinge begleiten, sprich, dass wir Journalisten informieren, dass da diese Aussendung kommt, und dass das eine erzählenswerte Geschichte ist und dass wir für Rückfragen zur Verfügung stehen, dass wir zuhören: Was braucht ihr, wenn ihr einen Artikel darüberschreibt? Also solche Dinge passieren, PR funktioniert auch über Relations [...]“ (Florian Novak von Tonio (Abs. 28)).

Weiterhin betonte Novak, dass Tonio von der Zusammenarbeit profitiert „solange wir in dem Zusammenhang erwähnt werden“ (Abs. 44) und es seinem Start-up hilft, „wenn ein etablierter Player sagt: Wir arbeiten mit denen zusammen“ (Abs. 44). Novaks Einsatz zahlte sich aus, denn kurz nach der Pressemitteilung wurde in den Medien über sein Start-up berichtet. Zu diesen Medien gehörte unter anderem die Wiener Tageszeitung derstandard.at und das österreichische Innovations-Portal trendingtopics.at.

Derstandard.at veröffentlichte so im Oktober 2017 online einen Artikel mit dem Titel „Wiener Start-up Tonio aktiviert für Turner den Second Screen“. Dieser teilte mit: „Dort laufe “Gumball VIP“, so der Name des interaktiven Zusatzangebots zum Fernsehcartoon, als Pilotprojekt, sagt Tonio-Gründer Florian Novak dem STANDARD: “Turner hat das Glück, dass sie im Kleinen Dinge ausprobieren können.“ Da sei weiterführend viel möglich, hofft Novak.““ (EQ# 22 der Standard, 2017). Auch trendingtopics.at berichtete von der Kooperation und erwähnte sogar, dass Novak selbst die Informationen zuspielte: „Wie Novak Trending Topics wissen ließ, ist er jetzt eine Partnerschaft mit der CNN-Mutter Turner Broadcasting System eingegangen. Die Tonio-Technologie wird ab Samstag bei Cartoon Network, einer Tochter von Turner, im Rahmen der Zeichentrick-Serie „Gumball VIP“ eingesetzt.“ (EQ# 24 trendingtopics, 2017).

Nach der Berichterstattung über „Gumball VIP“ Ende 2017 (EQ# 25 animationmagazine, 2017; EQ# 26 infodigital, 2017; EQ# 27 licensing.biz, 2017 uvm.), worüber das Start-up auch selbst auf seiner Website berichtete (EQ# 13 Tonio, 2018), wurde auch ein Jahr später über die Ausweitung von „Gumball VIP“ auf mehrere internationale Märkte in den Medien berichtet (EQ# 28 adobomagazine, 2018; EQ# 29 animationmagazine, 2018; EQ# 30 broadbandtvnews, 2018 uvm.). Auch TBS veröffentlichte diesen Erfolg auf seiner Website (EQ# 31 TBS, 2018, 2018) und broadbandtvnews.com zitierte den Chief Content Officer von Turner International, welcher über die „innovative technology“ berichtete, die der Organisation bei der Digitalisierung des linearen Fernsehens half (EQ# 30 broadbandtvnews, 2018).[1]

Narrativ: Technologie als Hilfsmittel für Fokus auf Kerngeschäft

Wie bereits unter 1.7 in Phase zwei beschrieben, nutzte Prokop in seiner Überzeugungsarbeit an interne Stakeholder bei TBS den Satz „Let’s agree on technology [...] but compete on content“ (EQ# 7 TBS, 2018, S. 16). Als Tonio nach dem erfolgreichen Pilotprojekt im Oktober 2018 seinen Newsletter verschickte, hieß es darin: „Creators should compete with their content, not their consumer data.“ (EQ# 21 Tonio, 2018). Diese beiden Aussagen sind sich sehr ähnlich und

[1] Abgesehen von der Legitimitätssteigerung für Tonio profitierte auch dessen Schlüsselkunde von dem Medienecho. Feldnotizen zeigen, dass über die Kooperation berichtende Artikel auch intern bei TBS geteilt wurden und speziell Prokop für seine Arbeit dadurch Lob erfuhr. Davon abgesehen führte die mediale Präsenz auch zu steigender Akzeptanz der Mitarbeiter für die innovative Technologie. Da die ersten Medienartikel von der Pressesendung des Medienunternehmens handelten, diese Artikel aber ohne die Bemühungen von Novak nicht in der Art erschienen wären und diese in der Konsequenz wiederum intern bei TBS Legitimitätssteigerungen und Lob für Prokop auslösten, zeigt sich dadurch erneut die Symbiose der Überzeugungsarbeit beider Organisationen.

stellen beide Tonios innovative Audio-Decoding Technologie als Hilfsmittel dar, nur, dass die ermöglichte Rückbesinnung aufs Kerngeschäft in Phase drei nicht mehr von Prokop an Mitarbeiter von TBS kommuniziert wurde, sondern Tonio eben dieses Narrativ nutzte, um neue potenzielle Kunden von sich und der innovativen Technologie zu überzeugen.

Eine solche Übernahme von Legitimationsmechanismen durch einen anderen Akteur, wie zum Beispiel bei der Innovations- und Wachstumsvision (siehe Phase 2: 1.5) oder der QR-Code Analogie (Siehe Phase 2: 1.10), zeigten die Daten bereits in Phase zwei, weshalb das Zusammenspiel von TBS und Tonio in dieser zweiten Phase erstmalig als symbiotisch charakterisiert wurde. Dass Tonio Prokops Narrativ der Technologie, welche wieder den Fokus auf das Kerngeschäft ermöglicht, aus Phase zwei in Phase drei weitertrug, macht klar, dass sich die Symbiose im Prozess sowohl auf die Legitimationsarbeit von TBS als auch auf die von Tonio, im Zuge des weiteren Wachstums, erstreckte. Dass Tonio dieses Narrativ übernahm, ist für die Feststellung des symbiotischen Zusammenspiels, welches den Prozess durchdrang, somit von großer Bedeutung.

Narrativ: Erfolgreiches Pilotprojekt

Die Daten lassen darauf schließen, dass die symbolische Handlung Novaks, die Medien aktiv auf die neueste Kooperation des Start-ups mit einem internationalen Medienunternehmen anzusetzen, kurz vor dem Pilotprojekt stattfand. Dessen Erfolg stand daher zu diesem Zeitpunkt noch nicht fest. Doch nachdem TBS die positive Beeinflussung des Engagements seiner Zuschauer und eine Steigerung der Reichweite durch das Pilotprojekt feststellte, integrierte Tonio diesen Erfolg in sein Narrativ, welches dem Start-up bei der Akquise weiterer Kunden half. Folgende Zitate von CEO Florian Novak zeigen diese präzise Legitimationsarbeit im Sinne von „Cultural Entrepreneurship“:

> „Also für uns ist natürlich die Referenz, dass ein großer Player mit uns Erfolg hat, den er ohne uns nicht hätte, ein Türöffner [...] und wenn wir es tun können, verweisen wir auch darauf [...] allein die Tatsache, dass wir sagen: Turner hat mit uns in den USA gearbeitet und es hat funktioniert, ist in Gesprächen in USA einfach eine comforting position. Oder sozusagen, ja, wenn man mit Konzernbären spricht, wo viele die Sorge entwickeln zu sagen: Wie schaue ich, dass mir was nicht zum Verhängnis wird, weil ich meinen Job behalten will und da kommt jemand und sagt: Du, das ist kein Experiment, sondern war schon in Einsatz, dann stabilisiert das den Akquisitionsprozess auch sehr.“ (Abs. 40).

> „Die Überraschung für uns [...] war, dass die Reichweiten gestiegen sind. Und das hätte ich nicht wagen getraut, aber das zeigt die Tangente zu sagen: Wenn ihr einen Beitrag leistet, dass ihr relevant am Smartphone seid, dann wird sich das positiv auf eure Reichweiten auswirken. Und das hilft uns natürlich im Moment auch bei weiteren Gesprächen schon sehr.“ (Abs. 38).

Novaks Aussage macht deutlich, dass Tonio nach dem Erfolg bei TBS versuchte, insbesondere Kunden in den USA zu akquirieren, wovon das Start-up auch in seinem Newsletter (EQ# 21 Tonio, 2018) berichtete. Doch auch die Zusammenarbeit mit TBS soll in Zukunft weiter ausgebaut werden, wie Prokop im Interview bestätigte: „Wir arbeiten weiter zusammen“ (Abs. 64).

4.4.2. Organisationale Mechanismen

Kooperation mit prestigevollem Kunden

Für ein Start-up wie Tonio, welches noch eher wenige Erfolge vorzuweisen hatte, war die Kooperation mit einem internationalen Medienunternehmen wie TBS eine wertvolle Legitimitäts-quelle, doch auch für TBS war die Zusammenarbeit mit einem innovativen Start-up wie Tonio nützlich, um Fortschrittlichkeit zu demonstrieren. Diese Interessen führten dazu, dass TBS Ende 2017, kurz bevor das Pilotprojekt „Gumball VIP“ startete, „in enger Abstimmung mit [Tonio]“ (Florian Novak von Tonio (Abs. 26)), deren Zusammenarbeit in einer Pressemitteilung öffentlich machte. Dies stellte generell einen organisationalen Mechanismus dar, weil dadurch organisationale Erfolge vermittelt wurden. Wie im vorherigen Punkt beschrieben, war diese Pressemitteilung die Grundlage für weitere Überzeugungsarbeit.

Um abschließend den symbiotischen Charakter der Überzeugungsarbeit übersichtlich darzu-stellen, sollen alle durchgeführten Überzeugungsmaßnahmen von Phase eins bis drei im Folgenden aufgelistet und diejenigen, welche auf Grund des Vorkommens in der vorherigen Phase die Symbiose aufzeigen, markiert werden.

5. Diskussion

Diese Arbeit wurde durch die Einnahme von zwei Perspektiven motiviert, um die Forschung an den Themen Entrepreneurship und Innovation durch den Ansatz des „Cultural Entrepreneur-ships“ zu erweitern und den dort bestehenden „entrepreneur-centric focus“ (Lounsbury et al., 2019b, S. 5) auf andere Akteure und Akteurskonstellationen auszuweiten (Garud et al., 2014; Gegenhuber und Naderer, 2018; Gehman und Soublière, 2017; Lounsbury et al., 2019b). Die erste Perspektive begriff „Cultural Entrepreneurship“ als gemeinsame Leistung mehrerer Akteure. Die zweite Perspektive erweiterte bestehende Bezugsgruppen um B-to-B Schlüsselkunden. Es wurde der Frage nachgegangen, wie eine neue Organisation B-to-B Schlüsselkunden überzeugt und mobilisiert mit dessen innovativer Technologie zu arbeiten und wie interne Dynamiken innerhalb des Schlüsselkunden diesen Prozess beeinflussen. Die in drei zeitlich aufeinanderfolgenden Phasen aufgebaute Analyse des Prozesses der Gewinnung des B-to-B Schlüsselkunden TBS durch Tonio und der Konsequenz daraus, führte zu den folgenden Ergebnissen.

In Phase eins (Tonio gewinnt TBS als Schlüsselkunden) agierte Tonio im Sinne von „Cultural Entrepreneurship“ als

Phase 1:

1. Identitätsmechanismen
   - 1.1 Symbolische Handlung: Elevator Pitch
   - 1.2 Symbolische Handlung: Anwendungsbeispiel
   - 1.3 Narrativ: Problematik aus Kundenperspektive
   - 1.4 Narrativ: Anwendungsbeispiel
   - 1.5 Narrativ: Innovations- und Wachstumsvision
   - 1.6 Narrativ: Etablierung als Standard
   - 1.7 Narrativ: Privacy Aspekt
   - 1.8 Narrativ: Tonio als magische Technologie
   - 1.9: Analogien: QR-Code und FedEX

2. Organisationale Mechanismen
   - 2.1 Struktur & Erfolge

3. Assoziative Mechanismen
   - 3.1 Beziehungen zur Elite
   - 3.2 Prestigevolle Organisationsmitglieder

Phase 2:

1. Identitätsmechanismen
   - 1.1 ***symbolische Handlung:*** ***Anwendungsbeispiele***
   - 1.2 symbolische Handlung: Adaption an TBS
   - 1.3 Narrativ: Adaption an TBS
   - 1.4 Narrativ: Erfolg des Pilotprojektes
   - 1.5 ***Narrativ:*** ***Innovations- und Wachstumsvision***
   - 1.6 Narrativ: Technologie als Hilfsmittel für Fokus auf Kerngeschäft
   - 1.7 ***Narrativ:*** ***Tonio als magische Technologie***
   - 1.8 ***Narrativ:*** ***Privacy Aspekt***
   - 1.9 ***Narrativ:*** ***Etablierung als Standard***
   - 1.10 ***Analogien:*** ***QR-Code***

Phase 3:

1. Identitätsmechanismen
   - 1.1 Symbolische Handlung: Mediennutzung
     → Interner Legitimationszuwachs auch bei TBS
   - 1.2 ***Narrativ:*** ***Technologie als Hilfsmittel für Fokus auf Kerngeschäft***
   - 1.3 Narrativ: Erfolgreiches Pilotprojekt

2. Organisationale Mechanismen
   - 2.1 Kooperation mit prestigevollem Kunden

Abbildung 1: Auflistung aller Legitimationsmechanismen nach Phasen
(aus vorheriger Phase übernommene Teile als Zeichen der symbiotischen Zusammenarbeit hervorgehoben)

geschickter kultureller Akteur (Aldrich und Fiol, 1994; Jennings et al., 2015; Lounsbury et al., 2019a; Lounsbury und Glynn, 2001; Zott und Huy, 2007) und setzte identitätsschaffende, assoziative und organisationale Legitimationsmechanismen ein, um TBS von sich und seiner innovativen Technologie zu überzeugen (Fisher et al., 2017).

In Phase zwei (Pilotprojekt bei TBS – Interne Allianzpartner schaffen für Tonio Akzeptanz) wurde ersichtlich, dass hauptsächlich Prokop, als Tonios interner Allianzpartner bei TBS, die Legitimationsarbeit in einer Teil-Entrepreneurship-Rolle intern weiterführte, wobei dieser hauptsächlich Narrative einsetzte, um seine internen Stakeholder von der

Innovation zu überzeugen (Lounsbury und Glynn, 2001). Diese Notwendigkeit der Überzeugung interner Stakeholder stimmt mit Erkenntnissen aus der Innovationsliteratur überein (Garud et al., 2013; Van de Ven, 1986). Es wurde deutlich, dass die Legitimationsarbeit Prokops auch von der vorangegangenen Legitimationsarbeit Tonios inspiriert wurde, was erstmals das Bild einer symbiotischen Zusammenarbeit zweier Organisationen zur Legitimation einer innovativen Technologie entstehen ließ.

Danach verdeutlichte Phase drei (Konsequenz des erfolgreichen Pilotprojektes), dass Tonio durch den Einsatz organisationaler und identitätsschaffender Legitimitätsmechanismen erneut „Cultural Entrepreneurship" Leistungen erbrachte, um sich bei neuen potenziellen Kunden zu legitimieren. Insbesondere die symbolische Handlung der gezielten Mediennutzung identifizierte das Start-up in dieser Phase als fokussierten kulturellen Akteur (Lounsbury und Glynn, 2001; Petkova et al., 2013; Zott und Huy, 2007). Die symbiotische Komponente der gemeinsamen Legitimationsarbeit erstreckte sich bis in die dritte Phase, da Tonio, ebenso wie TBS in Phase zwei, Argumente der anderen Organisation in seine Legitimationsarbeit übernahm.

Diese Ergebnisse demonstrieren die befähigende Kraft von Kultur im Bereich der Legitimation von Start-ups und innovativen digitalen Technologien (Lounsbury et al., 2019b). „Cultural Entrepreneurship" dabei als gemeinsame Aktivität mehrerer Akteure im Sinne von „Distributed Entrepreneurship" zu begreifen, steht aktuell noch am Anfang (Garud et al., 2014; Gegenhuber, T., 2017; Gegenhuber und Naderer, 2018; Lounsbury et al., 2019b), wird jedoch durch die in dieser Forschungsarbeit herausgearbeitete symbiotische Zusammenarbeit von Tonio als Start-up und TBS als dessen B-to-B Schlüsselkunde in einer Teil-Entrepreneurship-Rolle erweitert. Damit einher geht die gewonnene Erkenntnis, dass B-to-B Schlüsselkunden eine wichtige Bezugsgruppe für kulturelle Überzeugungsarbeit darstellen, welcher die einschlägige Literatur aktuell ebenso wenig Beachtung schenkt (Fisher et al., 2017; Hargadon und Douglas, 2001; Lounsbury und Glynn, 2001). Nachfolgend soll die symbiotische Zusammenarbeit von Start-ups und deren B-to-B Schlüsselkunden thematisiert und der theoretische Beitrag dieser Arbeit erörtert werden.

5.1. Innovationsprojekte als über Organisationsgrenzen hinweg fungierende Symbiosen

Im Kontext der Digitalisierung ist ein Medienunternehmen wie TBS dazu gezwungen, konkurrenzfähig zu bleiben. Eine Allianz mit dem Start-up Tonio und die mit diesem gemeinsam geleistete, kulturelle Überzeugungsarbeit für die Legitimation der innovativen Technologie, machte dies für TBS möglich. Eine solche Betrachtung von Innovationsprojekten, als über Organisationsgrenzen hinweg fungierenden Symbiosen, findet in der Literatur so jedoch noch nicht statt. In diesem Zusammenhang wurde nach dem Ressourcenabhängigkeitsansatz bisher lediglich von Allianzen und Kooptation gesprochen (Pfeffer & Salancik, 1978; in Wry et al., 2013; Santos und Eisenhardt, 2009). Hierbei wurden meist ausschließlich Entrepreneure in der aktiven Rolle gesehen, weshalb Kooptation in diesem Zusammenhang bedeutet, dass neue Organisationen etablierte Organisationen, beispielsweise durch Allianzen, an sich binden (Santos und Eisenhardt, 2009). In diesem Sinne hat Tonio durch die gezielte kulturelle Überzeugungsarbeit und die Allianz mit Bernhard Prokop, rund um das Pilotprojekt „Gumball VIP", zwar Teile von TBS kooptiert, doch das Konzept von Kooptation greift hier zu kurz, denn es lässt ein einseitiges Bild von Legitimationsarbeit und den dahinterstehenden Beweggründen entstehen.

Der Kooptationsgedanke vernachlässigt, dass auch etablierte Organisationen auf Grund von eigenen Interessen Legitimationsmotive besitzen und durch diese ebenso aktiv wie Entrepreneure werden können. Eben dies war bei Tonio und TBS der Fall, denn abgesehen von den drastischen Veränderungen in der Medienwelt, welche einst bewährte Geschäftsmodelle reifer Medien-organisationen in Frage stellen, hat Prokop seine Legitimationsarbeit für Tonio und dessen Ziel einer Standardisierung der Technologie auch teilweise als karriereförderlich angesehen und positionierte sich auch deswegen als aktiver Allianzpartner Tonios in der symbiotischen Zusammenarbeit.

Die Forschung von Garud et al. (2014) sowie Gegenhuber und Naderer (2018) zu „Distributed Entrepreneurship" greift diesen Symbioseaspekt auf. Letztere fanden heraus, dass Entrepreneure und deren Bezugsgruppe in dialogischen Prozessen interagieren, was letztlich zu Legitimitätssteigerungen führt. Obwohl die Fallstudie zu Tonio und TBS nur qualitative Indizien für Legitimität hervorbringt, bestätigen die gefundenen Ergebnisse entlang des dargestellten Prozesses die der Autoren. So haben TBS und Tonio beide die Rolle des kulturellen Akteurs, welcher die Innovation legitimiert, eingenommen, und sich so zusammen für Tonios Legitimation in einem „entrepreneurial process", in welchem „entrepreneurs and their audiences [interact]" (Lounsbury et al., 2019b, S. 5), eingesetzt, dessen Ergebnis eine Akzeptanzsteigerung und die interne Ausweitung des Pilotprojektes auf internationale Märkte war. Garud et al. (2014) beschreiben, dass durch die Interaktion von Entrepreneuren und Stakeholdern Narrative entstehen, welche teilweise auch durch die Stakeholder beeinflusst werden. Auch in der Allianzsituation im Fall von Tonio und TBS wurde die Grenze zwischen den beiden Organisationen beziehungsweise. Akteuren verwischt und es kam zu einer symbiotischen und organisationsübergreifenden Legitimationsarbeit, bei der sich alle Akteure proaktiv mit einbrachten.

Es ist fraglich, wie weit Tonio bei TBS gekommen wäre und wie das Start-up seinen weiteren Wachstumsweg beschreiten würde, wenn Prokop sich nicht intern beharrlich, auch durch den gezielten Einsatz von PowerPoint Präsentationen (Kaplan, 2011), für deren Innovation eingesetzt hätte und Tonio so teilweise auch zu weiterer kultureller Überzeugungsarbeit inspirierte (Garud et al., 2014).

Das in dieser Arbeit geschaffene Bild einer Symbiose stimmt mit den zu „Distributed Entrepreneurship" bestehen-

den Erkenntnissen (Garud et al., 2014; Gehenhuber & Naderer, 2018; Lounsbury et al., 2019b) überein und erweitert diese gleichzeitig. Die in dieser Arbeit gewonnenen Ergebnisse führen darüber hinaus dazu, den bestehenden Fokus auf den Entrepreneur, als alleinigen Erbringer kultureller Managementleistungen (Gehman und Soublière, 2017; Lounsbury und Glynn, 2001) auf B-to-B Schlüsselkunden auszuweiten, womit aktuellen Aufrufen aus der „Cultural Entrepreneurship“ Literatur nachgekommen wird (Lounsbury et al., 2019b).

An dieser Stelle soll verdeutlicht werden, was es bedeutet, „Distributed Entrepreneurship“ als Symbiose zu begreifen. Dafür ist es hilfreich, die drei verschiedenen Ebenen der Symbiose, wie sie im Prozess vorkamen, zu rekapitulieren. Ebene eins beinhaltete die Verknüpfung der Legitimationsarbeit verschiedener Akteure im Laufe eines Legitimationsprozesses einer Innovation. Diese Ebene zeigte sich darin, dass Prokop in Phase zwei zum kulturellen Akteur wurde und die Technologie durch den Einsatz von Identitätsmechanismen, wie der QR-Code Analogie, legitimierte, wobei einige dieser Mechanismen nicht von ihm selbst stammten. Viel mehr hatte er diverse „Cultural Entrepreneurship” Leistungen von Tonio aus Phase eins übernommen und setzte diese für die Überzeugung seiner internen Stakeholder ein. In der darauffolgenden Phase drei agierte Tonio identisch, indem Novak zur Akquise weiterer Kunden, ein zuvor von seinem Schlüsselkunden entwickeltes Narrativ übernahm (Let’s agree on technology but compente on content). Insbesondere diese zweite Stufe, in welcher das Start-up Legitimationsmechanismen seines B-to-B Schlüsselkunden nutzte und für weiteres Wachstum durch die Akquise neuer Kunden einsetzte, findet bisher in der Literatur nicht statt. Welche Mechanismen im Laufe des Prozesses von Tonio und TBS selbst entwickelt und welche übernommen wurden, zeigt nachfolgend ein Symbiosemodell (Abbildung 2).

Die zweite Ebene der Symbiose meint, dass TBS und auch Prokop als Einzelperson die Arbeit mit der innovativen Technologie vorantrieben und dessen Legitimität steigerten, weil der Erfolg der Technologie auch für sie selbst fruchtbar erschien und sie als kulturelle Entrepreneure motivierte. Eine dritte Ebene würdigt, dass bei TBS intern mit der Anzahl der auf Tonios Technologie basierenden erfolgreichen Projekte auch die Akzeptanz für die Technologie anstieg. Dies lässt den Gedanken von selbstverstärkenden Effekten zu, welche auf Erfolg basieren und als eigenständiger Akteur die Legitimation von Innovationen innerhalb reifer Organisationen vorantreiben.

Im Prinzip wird die Zusammenarbeit in dieser Arbeit als Symbiose bezeichnet, um die Verflechtung der Überzeugungsarbeit verschiedener Akteure im Sinne von „Cultural Entrepreneurship“, die Teil-Entrepreneurship-Rolle von TBS und den darüberstehenden eigenständigen Interessen an dem Erfolg der Legitimation beziehungsweise der Technologie als solcher, zu verdeutlichen. So war es die gemeinsame Leistung von Tonio und TBS, über den Großteil des analysierten Zeitraums hinweg, welche zur Legitimation der innovativen Audio-Decoding Technologie und der beteiligten Akteure beitrug und das Start-up in weiterem Wachstum beflügelte. Diese Sicht auf gemeinsame kulturelle Legitimationsarbeit kann die „Cultural Entrepreneurship“ Literatur und das daraus entstandene Konzept des „Distributed Entrepreneurships“ (Gegenhuber, T., 2017; Gegenhuber und Naderer, 2018; Gehman und Soublière, 2017; Lounsbury und Glynn, 2001; Lounsbury et al., 2019b) sowie die Literatur zur Legitimation neuer Organisationen Fisher et al., 2017; Zimmerman und Zeitz, 2002) voranbringen. Folgende Grafik verdeutlicht die symbiotische Legitimationsarbeit von Tonio und TBS über die drei Phasen hinweg, wobei das darin gezeigte Modell sich primär auf die zuvor als erste Symbioseebene beschriebenen, identitätsbasierten Legitimationsmechanismen bezieht und aus der Sicht Tonios formuliert wurde.

Eben dieses Prozessdenken ermöglicht einen weiteren Beitrag zur „Cultural Entrepreneur-ship“ Literatur, wenn diese mit der Literatur zu Innovationsmanagement (Garud et al., 2013; Van de Ven, 1986) in Verbindung gesetzt wird. Da in dieser Arbeit bei der Analyse der Prozess als Einheit gewählt wurde, zeigte sich, dass beide Literaturen im Kontext von Innovation und Entrepreneurship versuchen, ähnliche Probleme zu lösen, die Innovationsmanagement Literatur aber beispielsweise ein besseres Verständnis von internen Prozessen zu haben scheint. Dieses Verständnis kann insbesondere auf die Ereignisse in Phase zwei angewendet werden, in welcher sich Prokop innerhalb seiner reifen Organisation für die Akzeptanz der Innovation einsetzt. Vereinfacht beschrieben erklärt „Cultural Entrepreneurship“, wie ein Akteur einen anderen überzeugen kann. Innovationsmanagement hingegen geht stärker auf interne Problematiken bei der Einführung einer solchen Innovation ein.

Werden beide Ansätze auf Grund der Parallelen in den Ergebnissen in Zusammenhang gesetzt, könnten Phänomene im Bereich Innovation und Entrepreneurship umfassender durchdrungen werden. So hat Tonio die Bezugsgruppe der internen Stakeholder Fisher et al., 2017) innerhalb seines Schlüsselkunden lediglich durch die Allianz (Pfeffer & Salancik, 1978; in Wry et al., 2013; Santos und Eisenhardt, 2009) mit Prokop und durch dessen Innovationsmanagement, im Sinne eines Problembewusstseins für die Schwierigkeit der Vermittlung von Notwendigkeit oder Ideen der Technologie (Van de Ven, 1986) und einer entsprechenden Beharrlichkeit, überzeugen können. Zudem setzten sowohl Tonio als auch TBS Storytelling oder andere Narrative ein, um die innovative Technologie zu legitimieren (Lounsbury und Glynn, 2001).

Ein weiterer Beitrag, den diese Arbeit leistet, besteht in der Thematik der Kompetenzsteigerung, da Tonios Technologie TBS half, dessen Kerngeschäft zu digitalisieren. Lounsbury und Glynn (2001) beschrieben in ihrer Arbeit lediglich, dass Organisationen mit kompetenzfördernden Innovationen Unternehmensidentitäten schaffen sollten, die bestehende Branchenpraktiken ergänzen, ohne zu erklären, wie dies möglich ist. Hierfür liefert Tonios kulturelle Überzeugungsarbeit aus Phase eins das fehlende Beispiel und zeigte, dass unter anderem Pitches, Anwendungsbeispiele und Vi-

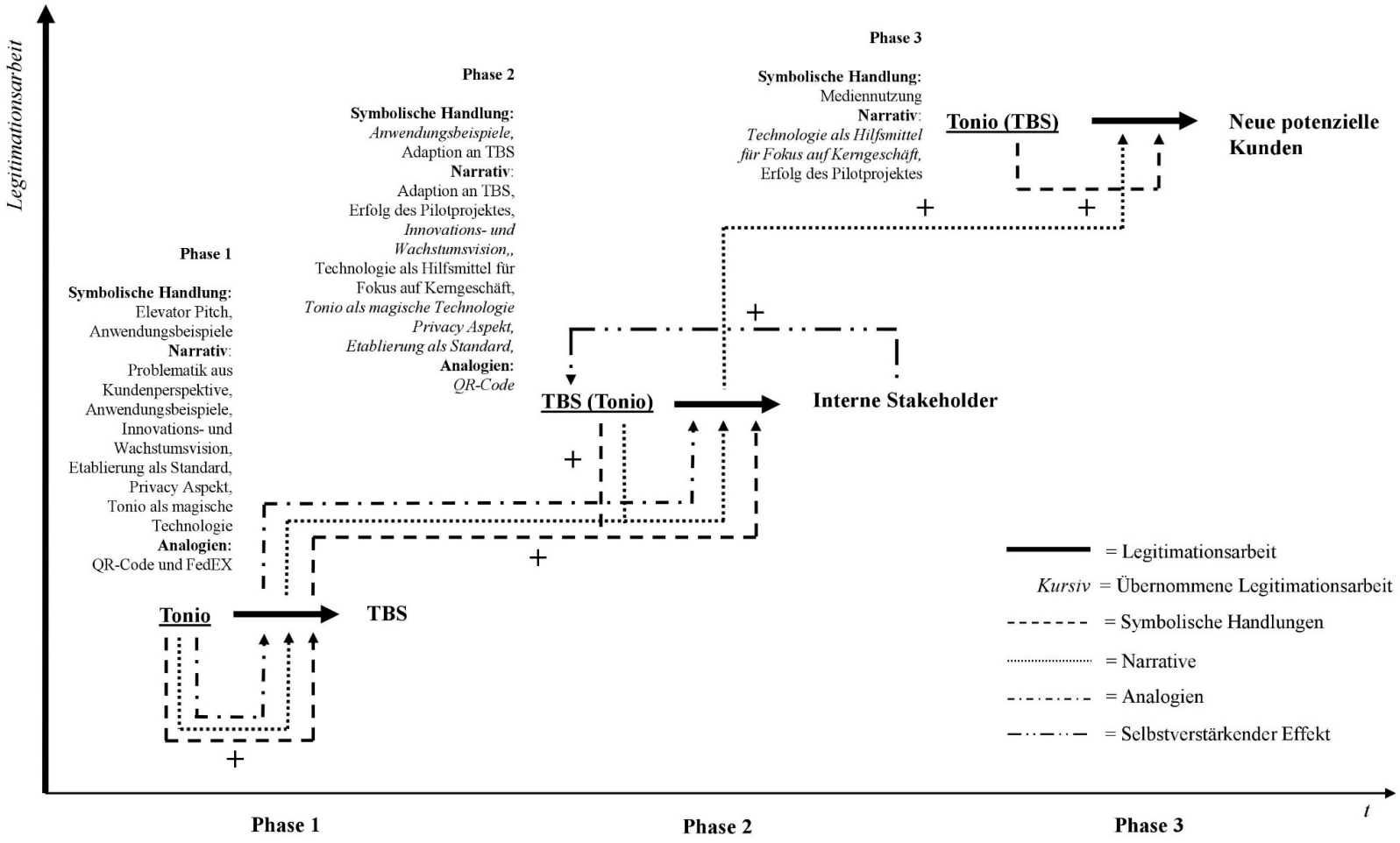


Abbildung 2: Symbiosemodell zur Legitimationsarbeit von Tonio und TBS

sionen, welche in der Branche bestehende Probleme lösen, kompetenzsteigernde Innovationen legitimieren können. Dadurch wird auch Tushman und Anderson (1986) sowie Abernathy und Clark (1985) ein Beispiel gegeben, wie eine von ihnen beschriebene kompetenzsteigernde Innovation im organisationalen Umfeld durch kulturelle Leistungen legitimiert werden kann und ergänzt deren Arbeit somit um eine kulturelle Komponente.

5.2. Ein möglicher Weg zur Legitimation bei B-to-B Schlüsselkunden

Bisherige Arbeiten zeigten lediglich, wie Entrepreneure die Bezugsgruppen der Geldgeber (Fisher et al., 2017) oder die der Wettbewerber, Kunden und Investoren (Lounsbury und Glynn, 2001) von sich überzeugen können. Die Ergebnisse dieser Arbeit hingegen zeigen eindeutig die Relevanz eines B-to-B Schlüsselkunden wie TBS für die Legitimation einer Innovation eines Start-ups, wodurch B-to-B Schlüsselkunden automatisch zu einer wichtigen Bezugsgruppe von „Cultural Entrepreneurship“ Leistungen solcher Start-ups werden. Denn bevor diese in einem symbiotischen Prozess gemeinsam im Sinne von „Distributed Entrepreneurship“ agieren, muss ein Start-up den potenziellen Kunden zunächst einmal von sich überzeugen (Phase 1). Dazu, wie ein Technologie Start-up wie Tonio, dies durch „Cultural Entrepreneurship“ Leistungen erreichen könnte, liefert diese Fallstudie eine erste mögliche Antwort.

Tonio setzte zu Beginn gezielt Identitätsmechanismen ein. Diese bestanden aus symbolischen Handlungen und Narrativen, wie zum Beispiel einem Elevator-Pitch, Anwendungsbeispielen, Analogien oder dem Aufzeigen einer ansprechenden, das unternehmerische Kapital und die eigenen Ziele beinhaltenden Vision (Hargadon und Douglas, 2001; Lounsbury und Glynn, 2001; Zott und Huy, 2007). Außerdem setzte Tonio organisationale und assoziative Mechanismen (Fisher et al., 2017) als kulturelle Leistungen ein, die TBS von der innovativen Technologie überzeugten und in der Konsequenz innovierenden Entrepreneuren eine Orientierung in der an B-to-B Schlüsselkunden gerichteten Legitimationsarbeit bieten können.

Die Beiträge, die diese Fallstudie zur bestehenden Literatur leistet, werden in nachfolgender Tabelle zusammenfassend den Phasen ein bis drei zugeordnet, wobei jedoch kein Anspruch auf Vollständigkeit erhoben wird.

6. Fazit und Ausblick auf künftige Forschung

Der Fall von Tonio und TBS zeigte, wie ein innovatives Start-up und dessen B-to-B Schlüsselkunde, in einer Organisationsgrenzen verwischenden Symbiose, gemeinsam „Cultural Entrepreneurship“ Leistungen für die Legitimation des Start-ups und dessen innovativer Technologie erbringen. Der in drei Phasen unterteilte Legitimationsprozess war dabei neben der gezielten kulturellen Überzeugungsarbeit durch das Start-up, sowohl von dem aktiven Einsatz interner Allianzpartner innerhalb des B-to-B Schlüsselkunden geprägt, als auch davon, dass über den gesamten Prozess hinweg, beide Organisationen Teile der Legitimationsarbeit der anderen Organisation, in Form von symbolischen Handlungen, Narrativen oder Analogien, in die eigene, an neue Bezugsgruppen gerichtete, Legitimationsarbeit übernahmen. Dieses Verständnis einer symbiotischen Zusammenarbeit mehrerer Akteure für die Legitimation einer Innovation erweitert die „Cultural Entrepreneurship“ Literatur (Fisher et al., 2017; Lounsbury und Glynn, 2001) und insbesondere das dort bestehende Konzept des „Distributed Entrepreneurships“ (Garud et al., 2014; Gegenhuber und Naderer, 2018; Lounsbury et al., 2019b). Zudem werden dadurch bestehende Erkenntnisse zu Allianzen und Kooptation (Santos und Eisenhardt, 2009) sowie zu Innovationsmanagement (Garud et al., 2013; Van de Ven, 1986) erweitert.

Abgesehen von den geleisteten Beiträgen, treffen auf diese Arbeit die üblichen Beschränkungen eines Fallstudienansatzes zu. Zum einen wurde eine Einzelfallstudie durchgeführt, weshalb es wichtig ist, sich den spezifischen Charakteristika aus dem Kontext von Tonio und TBS bewusst zu sein, bevor die Ergebnisse auf andere Start-ups und deren B-to-B Schlüsselkunden übertragen werden. Dies gilt insbesondere für das Symbiosemodell. Zum anderen basiert die Darstellung der Ereignisse hauptsächlich auf retrospektiven Interviews, weshalb das Risiko von verzerrten Interpretationen anerkannt und daraufhin bewusst Menschen interviewt wurden, die aktuell noch in die Arbeit mit der Audio-Decoding-Technologie involviert sind. Hinzu kommt, dass jeder dieser Menschen stark in die Technologie und die damit durchgeführten Projekte involviert ist und dementsprechend ein großes Interesse an der Darstellung einer Erfolgsgeschichte hat.

Diese Beschränkungen eröffnen jedoch gleichzeitig Möglichkeiten für zukünftige Forschung, um insbesondere das Verständnis einer symbiotischen Zusammenarbeit von Start-ups und B-to-B Schlüsselkunden, im Zuge der Legitimation von Innovationen und vor dem Hintergrund von „Cultural Entrepreneurship“ (Lounsbury und Glynn, 2001; Lounsbury et al., 2019b) und „Distributed Entrepreneurship“ (Garud et al., 2014; Gegenhuber und Naderer, 2018), zu erweitern. Zunächst könnte das aufgestellte Modell der symbiotischen Zusammenarbeit (Abbildung 2) im Innovations- und Entrepreneurshipkontext daher durch die Studie weiterer innovativer Start-ups und deren B-to-B Schlüsselkunden ausgebaut werden. Dies würde gleichzeitig die Relevanz von B-to-B Schlüsselkunden als Bezugsgruppe weiter herausstellen und so auch den praktischen Wert der Ergebnisse für innovierende Entrepreneure steigern (Fisher et al., 2017). Da andere Organisationen als Tonio und TBS, mit sehr hoher Wahrscheinlichkeit auf andere Art und Weise Legitimationsarbeit leisten, um deren Bezugsgruppen von Innovationen zu überzeugen, könnte untersucht werden, welche Ebenen bezeichnend für eine Symbiose sind. Sind es vielleicht gar nicht primär die übereinstimmenden Legitimationsmechanismen, sondern viel eher selbstverstärkende Effekte innerhalb des B-to-B Schlüsselkunden, oder ganz andere? Da die Feststellung solcher Effekte eine temporäre Dimension benötigt, könnte

Tabelle 4: Zusammenfassung der Beiträge aus dieser Fallstudie

	Phase 1	Phase 2	Phase 3
Literatur	Cultural Entrepreneurship (Lounsbury und Glynn, 2001) Symbolische Handlungen (Zott und Huy, 2007) Innovationsdesign (Hargadon und Douglas, 2001) Heterogene Bezugsgruppen kult. Überzeugungsarbeit (Fisher et al., 2017) Kompetenzsteigernde Innovationen (Abernathy und Clark, 1985; Tushman und Anderson, 1986)	Cultural Entrepreneurship (Lounsbury und Glynn, 2001) Symbolische Handlungen (Zott und Huy, 2007) Innovationsdesign (Hargadon und Douglas, 2001) Heterogene Bezugsgruppen kult. Überzeugungsarbeit (Fisher et al., 2017) Allianzen und Kooptation (Santos und Eisenhardt, 2009) „Distributed Entrepreneurship“ (Garud et al., 2014; Gegenhuber und Naderer, 2018; Gegenhuber, T., 2017; Lounsbury et al., 2019b) Klassisches Innovationamanagement (Garud et al., 2013; Van de Ven, 1986)	Cultural Entrepreneurship (Lounsbury und Glynn, 2001) Symbolisches Management (Zott und Huy, 2007)
Fallstudie	Tonio hat nach der „Cultural Entrepreneurship“ Literatur als kultureller Akteur gehandelt	Allianz mit B-to-B Schlüsselkunde, Symbiotische Zusammenarbeit (auf drei Ebenen)	Symbiotische Zusammenarbeit (Tonio übernimmt Legitimitäts-argumente von TBS)
Beitrag	Bestätigung und Präzision der bereits bestehenden Erkenntnisse in der Literatur. Insbesondere Präzision von Legitimationsarbeit für kompetenzsteigernde Innovationen sowie B-to-B Schlüsselkunden als relevante Bezugsgruppe.	Erweiterung von „Cultural Entrepreneurship“ und „Distributed Entrepreneurship“ sowie Kooptation im Sinne einer über Organisationsgrenzen hinweg fungierenden symbiotischen Zusammenarbeit mehrerer Akteure im Legitimationsprozess von Innovationen Erweiterung der „Cultural Entrepreneurship“ und „Distributed Entrepreneurship“ Literatur um die Übernahme von Legitimationsargumenten durch Start-ups von B-to-B Schlüsselkunden für weiteres Wachstum Erweiterung der „Cultural Entrepreneurship“ Literatur durch Aufnahme von B-to-B Schlüsselkunden in relevante Akteurs- und Bezugsgruppe bezüglich kultureller Überzeugungsarbeit Erweiterung der klassischen Innovationsmanagementliteratur um Symbiose sowie Verbindung von „Cultural Entrepreneurship“ und Innovationsmanagement im Legitimationsprozess	

die Idee der Symbiose zudem innerhalb eines längeren Zeitraums, eventuell im Rahmen einer Verlaufsstudie, untersucht werden.

Hierbei könnten insbesondere mögliche negative Effekte von B-to-B Schlüsselkunden auf Start-ups beleuchtet werden, denn im Falle der Symbiose von Tonio und TBS wirkte sich ein B-to-B Schlüsselkunde ausgesprochen positiv auf ein wachsendes Start-up aus, doch aufbauend auf Fisher, Kotha und Lahiri (2016) könnte ebenso erforscht werden, ob beziehungsweise ab welchem Zeitpunkt B-to-B Schlüsselkunden Start-ups vielleicht sogar auf ihrem Wachstumspfad behindern. Zudem hat in dem hier untersuchten Fall ein Start-up einen Schlüsselkunden eingenommen (Santos und Eisenhardt, 2009). Doch auch die Frage danach, wie sich eine symbiotische Zusammenarbeit auf das Verhalten von Organisationen auswirkt, bei welcher das Start-up von einem B-to-B Schlüsselkunden eingenommen und dadurch in ein Abhängigkeitsverhältnis, wie es auch der Ressourcenabhängigkeitsansatz (Wry et al., 2013) beschreibt, gebracht wird, könnte dem in dieser Arbeit untersuchten Beispiel einer fruchtbaren Symbiose mögliche schädliche Auswirkungen von Symbiosen gegenüberstellen. In diesem Sinne laden die gewonnenen

Ergebnisse dazu ein, weitere Fälle zu untersuchen, in welchen mehrere Akteure gemeinsam „Cultural Entrepreneurship“ Leistungen erbringen, um mögliche Nachteile von Symbiosen sowie das volle Potenzial symbiotischer Kooperationsprozesse im Kontext von Innovation, Entrepreneurship und Kultur für die Legitimation von Innovationen auszuschöpfen und weitere Effekte von Symbiosen in anderen Feldern oder Branchen zu erforschen.

Literatur

Abernathy, W. J. und Clark, K. B. Innovation: Mapping the winds of creative destruction. *Research policy*, 14(1):3–22, 1985.

Aldrich, H. E. und Fiol, C. M. Fools rush in? the institutional context of industry creation. *Academy of Management Review*, 19(4):645–670, 1994.

Edmondson, A. C. und McManus, S. E. Methodological fit in management field research. *Academy of Management Review*, 32(4):1246–1264, 2007.

EQ# 10 TBS. Gumball VIP Activity Performance, 2018.

EQ# 11 TBS. Executive Summary AT&T and Tonio, 2018.

EQ# 12 TBS. Gumball Info for Each Local Market, 2018.

EQ# 13 Tonio, 2018. URL https://www.tonio.com/.

EQ# 14 Die Presse, 2018. URL https://diepresse.com/home/kultur/medien/5077247/ORF-holt-sich-SecondScreenHilfe-von-Tonio.

EQ# 15 PwC, 2018. URL https://www.pwc.de/de/technologie-medien-und-telekommunikation/gemo-2018.pdf.

EQ# 16 Deloitte, 2018. URL https://www2.deloitte.com/tr/en/pages/technology-media-and-telecommunications/articles/media-and-entertainment-industry-outlook-trends0.html.

EQ# 17 W&V, 2018. URL https://www.wuv.de/medien/so_will_lineares_fernsehen_ueberleben.

EQ# 18 TBS, 2019. URL https://www.turner.com/about-us.

EQ# 19 Radioszene, 2005. URL https://www.radioszene.de/news/loungefm_200605.htm.

EQ# 20 Ashoka, 2016. URL http://ashoka-cee.org/austria/ashoka-partner/wolfgang-neubert/.

EQ# 21 Tonio. Newsletter, 2018.

EQ# 22 der Standard, 2017. URL https://derstandard.at/2000066316109/Wiener-Startup-Tonio-aktiviert-fuer-Turner-den-Second-Screen.

EQ# 23 der Standard, 2018. URL https://derstandard.at/2000090265368/Turner-setzt-Tonio-App-aus-Oesterreich-in-zehn-weiteren-Laendern.

EQ# 24 trendingtopics, 2017. URL https://www.trendingtopics.at/tonio-kooperation-turner/.

EQ# 25 animationmagazine, 2017. URL http://www.animationmagazine.net/tv/turner-emea-and-tonio-team-for-gumball-vip-uk-experiment/.

EQ# 26 infodigital, 2017. URL https://www.infosat.de/digital-tv/gumball-vip-turner-emea-und-startup-tonio-starten-tv-experiment.

EQ# 27 licensing.biz, 2017. URL https://www.licensing.biz/entertainment/turner-and-tonio-partner-for-the-amazing-world-of-gumball-live-quiz-experiment.

EQ# 28 adobomagazine, 2018. URL https://adobomagazine.com/digital-news/icymi-gumball-vip-launch-singapore-philippines-australia-and-new-zealand.

EQ# 29 animationmagazine, 2018. URL http://www.animationmagazine.net/gaming/turners-gumball-vip-on-the-roll-internationally/.

EQ# 30 broadbandtvnews, 2018. URL https://www.broadbandtvnews.com/2018/10/29/turner-to-rollout-second-screen-gumball-internationally/.

EQ# 31 TBS, 2018. URL https://www.turner.com/pressroom/turner-international-announces-roll-out-gumball-vip-following-successful-uk-trial.

EQ# 5 TBS. Presentation Tonio TI_EC_Meeting, 2016.

EQ# 6 TBS. Strategy for Market Implementation, 2017.

EQ# 7 TBS. CNN Tonio Introduction_ print, 2018.

EQ# 8 TBS. Intro Tonio, 2018.

EQ# 9 TBS. Gumball VIP What Is It Intro Presentation, 2018.

Fisher, G., Kuratko, D. F., Bloodgood, J. M., und Hornsby, J. S. Legitimate to whom? the challenge of audience diversity and new venture legitimacy. *Journal of Business Venturing*, 32(1):52–71, 2017.

Flick, U. Qualitative Sozialforschung. Eine Einführung (1. Aufl. der Neuausg.). Reinbek bei Hamburg: Rowohlt-Taschenbuch-Verlag, 2007.

Garud, R., Tuertscher, P., und Van de Ven, A. H. Perspectives on innovation processes. *Academy of Management Annals*, 7(1):775–819, 2013.

Garud, R., Gehman, J., und Giuliani, A. P. Contextualizing entrepreneurial innovation: A narrative perspective. *Research Policy*, 43(7):1177–1188, 2014.

Gegenhuber, T. und Naderer, S. When the petting zoo spawns into monsters: open dialogue and a venture's legitimacy quest in crowdfunding. *Innovation*, 19(857):1–36, 2018.

Gegenhuber, T. Cultural Entrepreneurship: Making audiences attend, understand, and value. Dissertation, Johannes Kepler Universität Linz, 2017.

Gehman, J. und Höllerer, M. A. Venturing into the cultural future: research opportunities at the nexus of institutions, innovation and impact. *Innovation*, 37(6):1–20, 2018.

Gehman, J. und Soublière, J.-F. Cultural entrepreneurship: from making culture to cultural making. *Innovation*, 19(1):61–73, 2017.

Hargadon, A. B. und Douglas, Y. When innovations meet institutions: Edison and the design of the electric light. *Administrative Science Quarterly*, 46 (3):476–501, 2001.

Helfferich, C. *Die Qualität qualitativer Daten. Manual für die Durchführung qualitativer Interviews*. (4. Auflage). Wiesbaden: Springer VS., 2011.

Jennings, J. E., Edwards, T., Jennings, P. D., und Delbridge, R. Emotional arousal and entrepreneurial outcomes: Combining qualitative methods to elaborate theory. *Journal of Business Venturing*, 30(1):113–130, 2015.

Kaiser, R. *Qualitative Experteninterviews: Konzeptionelle Grundlagen und praktische Durchführung*. Springer-Verlag, 2014.

Kaplan, S. Strategy and powerpoint: An inquiry into the epistemic culture and machinery of strategy making. *Organization Science*, 22(2):320–346, 2011.

Lamnek, S. und Krell, C. Qualitative Sozialforschung. Weinheim: Beltz. 2010.

Lounsbury, M. und Glynn, M. A. Cultural entrepreneurship: Stories, legitimacy, and the acquisition of resources. *Strategic Management Journal*, 22 (6-7):545–564, 2001.

Lounsbury, M., Cornelissen, J., Granqvist, N., und Grodal, S. Culture, innovation and entrepreneurship. *Innovation*, 21(1):1–12, 2019a.

Lounsbury, M., Gehman, J., und Ann Glynn, M. Beyond homo entrepreneurus: Judgment and the theory of cultural entrepreneurship. *Journal of Management Studies*, 2019b.

Mayring, P. Qualitative Inhaltsanalyse. Grundlagen und Techniken (12., überarbeitete Auflage). Weinheim: Beltz Verlag, 2015.

Nambisan, S., Lyytinen, K., Majchrzak, A., und Song, M. Digital innovation management: Reinventing innovation management research in a digital world. *Mis Quarterly*, 41(1), 2017.

Perez, L., Whitelock, J., und Florin, J. Learning about customers: Managing b2b alliances between small technology startups and industry leaders. *European Journal of Marketing*, 47(3/4):431–462, 2013.

Petkova, A. P., Rindova, V. P., und Gupta, A. K. No news is bad news: Sensegiving activities, media attention, and venture capital funding of new technology organizations. *Organization Science*, 24(3):865–888, 2013.

Rao, H. The social construction of reputation: Certification contests, legitimation, and the survival of organizations in the american automobile industry: 1895–1912. *Strategic Management Journal*, 15(S1):29–44, 1994.

Santos, F. M. und Eisenhardt, K. M. Constructing markets and shaping boundaries: Entrepreneurial power in nascent fields. *Academy of Management Journal*, 52(4):643–671, 2009.

Tushman, M. L. und Anderson, P. Technological discontinuities and organizational environments. *Administrative Science Quarterly*, pages 439–465, 1986.

Van de Ven, A. H. Central problems in the management of innovation. *Management Science*, 32(5):590–607, 1986.

Wry, T., Cobb, J. A., und Aldrich, H. E. More than a metaphor: Assessing the historical legacy of resource dependence and its contemporary promise as a theory of environmental complexity. *The Academy of Management Annals*, 7(1):441–488, 2013.

Yin, R. K. Case study research. Design and methods (Vol. 5, 4. ed.). Los Angeles: Sage, 2009.

Zimmerman, M. A. und Zeitz, G. J. Beyond survival: Achieving new venture growth by building legitimacy. *Academy of Management Review*, 27(3): 414–431, 2002.

Zott, C. und Huy, Q. N. How entrepreneurs use symbolic management to acquire resources. *Administrative Science Quarterly*, 52(1):70–105, 2007.

Junior Management Science 4(3) (2019) 460-477

Junior Management Science

journal homepage: www.jums.academy

Das Management von Produktrückrufen: Einflussfaktoren auf die Rückholung von Verbraucherprodukten

Annalena Düker

Freie Universität Berlin

Abstract

Immer mehr Produkte werden wegen sicherheitsrelevanter Mängel zurückgerufen. Doch trotz größter unternehmerischer Bemühungen, die eigenen Produkte mit einem möglichst großen Sicherheits- und Qualitätsanspruch in den Verkehr zu bringen, sind Fehler nie vollständig auszuschließen. Produktrückrufe gehören zur geschäftlichen Realität. Tritt der Ernstfall eines Produktrückrufes ein, geht es in erster Linie um die Sicherheit der Konsumenten, aber auch um immensen Reputationsschaden und hohen Aufwendungen für das betroffene Unternehmen.

Ziel der vorliegenden wissenschaftlichen Arbeit ist es, den Prozess eines erfolgreichen Produktrückrufes zu untersuchen und Maßnahmen abzuleiten, die ein Unternehmen präventiv und während eines Rückrufes ergreifen muss um alle betroffenen Produkte erfolgreich aus dem Markt zu entfernen. Basierend auf Erkenntnissen der aktuellen Produktrückrufforschung und der einschlägigen Gesetzgebung werden organisatorische, rechtliche sowie verhaltenswissenschaftliche Einflussfaktoren identifiziert, die berücksichtigt werden müssen, um eine vollständige Rückholung aller Produkte aus dem Markt zu gewährleisten.

Keywords: Produktrückruf; Rückrufmanagement; Krisenmanagement; Konsumentenverhalten; Produktkrise.

1. Einleitung

Explodierende Handy Akkus bei Samsung, keimbelastete Seifen von DM, Explosionsgefahr von Gasherden bei BOSCH, manipulierte Dieselautos in der Automobilindustrie, Plastikteile in Schokoriegeln bei Mars. Diese Aufzählung von Produktrückrufen aus dem Jahr 2016 ließe sich beliebig fortsetzen. Immer mehr Produkte werden wegen sicherheitsrelevanter Mängel zurückgerufen. Die Tendenz steigend (vgl. BAuA, 2017a). Lebensmittel, Verbraucherprodukte, Autos - die Bandbreite der betroffenen Güter reicht dabei quer durch das verfügbare Konsumangebot. Die Gründe für einen Anstieg von Produktrückrufen sind so vielfältig wie die Industrien, die von Rückrufen betroffen sind. So können Gründe in der gestiegenen Komplexität von Produkten gesehen werden, in höheren Entwicklungszeiten aufgrund einer steigenden Wettbewerbsintensität oder in einer globalen Arbeitsteilung infolge von Wertschöpfungsverlagerungen (vgl. Van Venrooy, 2015, S.1).

Kein Unternehmen möchte Produkte zurückrufen. Doch trotz aller Bemühungen, die eigenen Produkte mit einem möglichst großen Sicherheits- und Qualitätsanspruch in den Verkehr zu bringen, sind Fehler dennoch nie vollständig auszuschließen (vgl. Klindt und Wende, 2014, S.1). Umso wichtiger erscheint die Notwendigkeit für Unternehmen sich mit der Umsetzung von Produktrückrufen auseinanderzusetzen um im Ernstfall Konsumenten vor den Auswirkungen solcher Fehler erfolgreich zu schützen. Auch aufgrund der Tatsache, dass die Produktsicherheit in den letzten Jahren weltweit in den Vordergrund rückt und sich besonders in der Europäischen Union die gesetzliche und gesellschaftliche Relevanz eines starken Verbraucherschutzes erhöht, wird auch eine Auseinandersetzung mit dem Thema Produktrückruf immer wichtiger (vgl. Van Venrooy, 2015, S.5).

Vor diesem Hintergrund ist das Ziel der vorliegenden Arbeit, den Prozess eines erfolgreichen Produktrückrufes zu untersuchen und herauszuarbeiten, auf welchem Wege ein Unternehmen eine vollständige Rückholung aller unsicheren Produkte aus dem Markt erreichen kann, um die Sicherheit ihrer Konsumenten sicherzustellen und Schäden für das Unternehmen zu vermeiden.

Wie aufgezeigt sind alle Produkte, Industrien und Länder vom Risiko einer Rückrufsituation betroffen. Die Behandlung der Forschungsfrage unter Berücksichtigung aller gesetzlichen, länder- und produktspezifischen Gesichtspunkten erscheint wegen der unüberschaubaren Menge nicht zielfüh-

DOI: https://doi.org/10.5282/jums/v4i3pp460-477

rend. Aufgrund der starken wirtschaftlichen Verflechtungen zwischen Deutschland und den USA sollen sich die Untersuchungen der vorliegenden Arbeit auf Rückrufe von Verbraucherprodukten in der deutschen- und US-amerikanischen Rechtsordnung beschränken.

2. Der Produktrückruf

Zunächst sollen einige Grundlagen zum Produktrückruf aufgearbeitet werden um ein Grundverständnis des Themengebiets der vorliegenden Forschungsfrage zu erwirken.

2.1. Definition und Abgrenzung

Das Produktsicherheitsgesetz (ProdSG), als eine der entscheidenden gesetzlichen Regelungen des Produktrückrufs in der deutschen Rechtsordnung, definiert den Produktrückruf als „ jede Maßnahme, die darauf abzielt, die Rückgabe eines dem Endverbraucher bereitgestellten Produkts zu erwirken“ (§2 S. 25 ProdSG). Somit beschreibt das Gesetz den Produktrückruf als Vorgang des physischen Rückholens eines Produkts aus den Händen des Verbrauchers. Der Produktrückruf wird vom ProdSG deutlich von einer Produktrücknahme abgegrenzt. Der Unterschied liegt hierbei im Aufenthaltsort der unsicheren Produkte. Während sich bei einem Rückruf die betroffenen Produkte bereits bei den Verbrauchern befinden, ist eine Rücknahme dann erforderlich, wenn sich die Produkte noch in der Lieferkette befinden und den Endverbraucher noch nicht erreicht haben (vgl. §2 Nr.24 ProdSG).

Formen des Produktrückrufs

Rückrufe dienen dazu, die vom Hersteller bereits ausgelieferten betroffenen Produkte vorübergehend oder endgültig aus dem Verkehr zu ziehen (vgl. Tamme, 1996, S. 6). Die konkrete Ausgestaltung des Rückrufs erweist sich dabei als situationsabhängig. Rückrufe können sich hinsichtlich ihres beabsichtigten Zwecks, der Art der Entscheidungsfindung und des Grades der Öffentlichkeit voneinander unterscheiden. (vgl. Rettenbeck, 1994, S.16).

Hinsichtlich der Art der Entscheidungsbildung lässt sich zwischen einem freiwilligen und einem angeordneten Rückruf differenzieren. Ruft der Hersteller Produkte aus eigenem Entschluss zurück handelt es sich um einen freiwilligen Rückruf. Hierbei ist nicht entscheidend, ob der Hersteller sich erst dann für einen Rückruf entscheidet, nachdem beispielsweise Verbraucherverbände oder staatliche Stellen erheblichen Druck auf ihn ausgeübt haben. Entscheidend ist, dass der Rückruf auf der eigenen Entscheidung des Herstellers beruht (vgl. Tamme, 1996, S. 8). Wird der Rückruf nicht vom Hersteller entschieden, sondern von einer hoheitlichen Stelle rechtsverbindlich angeordnet, handelt es sich um einen angeordneten Rückruf (vgl. Pepels, 2016, S. 179). Hinsichtlich des verfolgten Zwecks kann zwischen einem vollständigen und einem vorübergehenden Produktrückruf unterschieden werden (vgl. Rettenbeck, 1994, S. 16). Produkte, deren Fehler nicht mehr behebbar sind, werden in einem vollständigen Rückruf endgültig vernichtet. Wird das unsichere Produkt hingegen nur überprüft und die Gefahrenquelle durch eine Reparatur beseitigt handelt es sich um einen vorübergehenden Rückruf, da das Produkt nur temporär „aus dem Verkehr“ gezogen wird (vgl. ebd.). Ein für die Forschungsfrage sehr relevantes Unterscheidungsmerkmal ist die Öffentlichkeit des Rückrufes. In einem öffentlichen Rückruf wendet sich der Hersteller über die allgemeinen Massenmedien an die Besitzer der unsicheren Produkte (vgl. Tamme, 1996, S. 10). Besteht der Fall, dass die Endverbraucher dem Hersteller nicht bekannt sind ist der öffentliche Rückruf meist die einzige Möglichkeit die Produktbenutzer zu erreichen (vgl. ebd.). Dagegen liegt ein stiller Rückruf vor, wenn der Hersteller keinen Gebrauch von den allgemeinen Massenmedien macht um die betroffenen Produktbesitzer zu erreichen. Dies trifft dann zu, wenn die Endverbraucher dem Hersteller bekannt sind und diese persönlich über Produktunsicherheiten informiert werden können (vgl. Tamme, 1996, S.9).

2.2. Auslöser für einen Produktrückruf

Es kann verschiedene Gründe geben, warum ein Unternehmen ein Produkt vom Markt nehmen muss oder sich aus eigener Entscheidung für einen Rückruf seiner Produkte entscheidet.

Sicherheitsprobleme

Produkte werden zurückgerufen, wenn Mängel die Sicherheit des im Markt befindlichen Produktes beeinträchtigen und somit die Produktverbraucher in ihrer Gesundheit oder Sicherheit gefährden (vgl. Eilert et al., 2017, S.113). Solch ein Produktversagen kann durch verschiedene Situationen hervorgerufen werden. Wurde bereits in der Entwicklungsphase eines Produktes nicht die erforderliche Sorgfalt aufgewendet, können Konstruktionsfehler entstehen, die die gesamte Produktionsserie betreffen (vgl. Lach und Polly, 2015, S.67). Ein Produktversagen kann ebenfalls durch Fabrikationsfehler ausgelöst werden, welche während der Produktion entstehen und bei Qualitätskontrollen nicht entdeckt wurden. Diese Fehlerart ist sehr problematisch, da sie nur einzelne Exemplare einer Produktionsserie betreffen kann und es schwierig zu bestimmten ist, welche genauen Produkte den Fehler aufweisen (vgl. Bodewig, 1999, S.179 f.). Ein Produktversagen kann auch dann gegeben sein, wenn ein Produkt fehlerfrei produziert wird, es jedoch zu einem unvorhersehbaren Fehlgebrauch durch Konsumenten kommt (vgl. Eilert et al., 2017, S. 113). Wenn ein Produktfehler eine ernsthafte Gefahr für Verbraucher zu Folge hat, ist ein Unternehmen gesetzlich verpflichtet einen Rückruf zu initiieren. Unter diesen Umständen kann eine Rückrufaktion ebenfalls von einer Behörde angeordnet werden (vgl. §26 Abs.4 ProdSG). Auch externe Faktoren können einen Einfluss auf die Entscheidung für einen Rückruf infolge von Sicherheitsproblemen haben. Werden Fehler am Produkt festgestellt, die kein ernstes Risiko darstellen, können sich Unternehmen trotz fehlender gesetzlicher Verpflichtung für einen Rückruf entscheiden, beispielsweise wenn eine besondere Sensibilität der Medien und der Gesellschaft über das betroffene Produkt besteht (z.B. bei Gefahren für Kinder) (vgl. Klindt und Wende, 2014, S.79).

Nichterfüllung der unternehmensinternen Qualitätsstandards und normative Gründe

Darüber hinaus kann sich ein Unternehmen aus eigener Entscheidung für einen Produktrückruf entscheiden, wenn zwar vom Produkt keine sicherheitsrelevante Gefährdung für den Konsumenten ausgeht, dieses jedoch nicht den eigenen unternehmensinternen Qualitätsstandards entspricht. Die Qualität eines Produktes ist ein entscheidendes Merkmal, damit das Produkt langfristig erfolgreich am Markt positioniert werden kann. Um den wachsenden Qualitätsanforderungen der Konsumenten gerecht zu werden, muss die Erwartungshaltung der Kunden erfüllt werden. Somit kann ein Produktrückruf bei aufkommenden Produktschwächen notwendig werden, um die Kritik von Konsumenten und daraus resultierende langfristige Imageschäden zu vermeiden (vgl. Van Venrooy, 2015, S. 29).

Auch ohne eine eigene Unternehmensentscheidung kann ein Rückruf durch eine behördliche Anordnung ausgelöst werden. Ein Unternehmen kann zu einem Rückruf gesetzlich verpflichtet werden, wenn von einem Produkt eine erhebliche Gefahr ausgeht (vgl. §26 Abs.4 ProdSG). Ebenfalls kann ein Rückruf auch bei Urheberrechtsschutz- und Wettbewerbsverletzungen gesetzlich angeordnet werden. So verpflichtete das Oberlandesgericht Düsseldorf 2017 ein Unternehmen zum Produktrückruf, nachdem dieses von einem Konkurrenten für die Verwendung des Werbespruchs „40% mehr Wirksamkeit auf Luftbefeuchtern“ abgemahnt wurde (Oberlandesgericht Düsseldorf, 2015, S.1).

2.3. Konsequenzen eines Rückrufs

Produktrückrufe können signifikante lang- und kurzfristige Konsequenzen für die Konsumenten und für betroffene Unternehmen hervorrufen. Zu denken ist insbesondere an Schäden bei den betroffenen Endkunden, die von leichten Verletzungen bis zu erheblichen Gesundheitsschädigungen oder sogar zum Tod reichen können (vgl. Tamme, 1996, S.7). Für das betroffene Unternehmen kann ein Rückruf sowohl direkte als auch indirekte Kosten verursachen. Die direkten Kosten beinhalten alle Ausgaben, die im direkten Zusammenhang mit dem Produktrückruf stehen. Hier lassen sich Aufwendungen für die Kompensation (remedy) nennen, wie Reparaturleistungen, vollständige oder teilweise Kaufpreisrückerstattungen, sowie alle weiteren Kosten, die mit der Produktrückholung verbunden sind (vgl. Raithel und Hock, 2017, S.1). Kommt es zu Verletzungen oder Todesfällen bei betroffenen Verbrauchern können für Unternehmen erhebliche Aufwendungen für Rechtsstreitigkeiten sowie daraus resultierende Schadensersatzzahlungen entstehen, welche ebenfalls zu den direkten Kosten gezählt werden (Eilert et al., 2017, S. 113). Die tatsächliche Größenordnung der direkten Kosten hängt von der Höhe der vom Rückruf betroffenen Produkte sowie der Anzahl der erfolgreich vom Markt entfernten defekten Produkte ab (vgl. ebd.). Ferner verursacht ein Rückruf auch indirekte Kosten. Diese zeigen sich in langfristigen Reputationsschäden durch Verluste an Vertrauen und Glaubwürdigkeit, langfristigen Kundenabwendungen, einer negativen Entwicklung des Aktienkurses, Umsatzeinbrüchen und einer Beeinträchtigung des Geschäftsverlaufs (vgl. Glasl und Klindt, 2012, S.14).

Bezugnehmend auf die Forschungsfrage wird durch eine Rückholung aller betroffener Produkte aus dem Markt die Gefahr für Verbraucher am effektivsten beseitigt, erhöht aber in der Regel auch die direkten Kosten eines Rückrufs (vgl. Bodewig, 1999, S.12). So könnte man annehmen, das Unternehmen aus einer kurzfristigen Kostenperspektive durchaus dazu motiviert sein könnten, den Rückruf unwirksam umzusetzen, um die direkten Kosten möglichst minimal zu halten. Es zeigt sich jedoch, dass Rückrufe, die von Unternehmen unpassend und wenig akkurat durchgeführt werden, sich zu ernsthaften Bedrohungen für Konsumenten, aber auch für das Unternehmen , ausweiten können (vgl. Van Venrooy, 2015, S.2, Gibson, 1995, S.230). Dagegen kann ein offensiver und kulanter Umgang mit Produktmängeln im Rahmen eines Rückrufs die langfristigen Schäden für Konsumenten und Unternehmen nicht nur begrenzen, sondern im Einzelfall die Unternehmensreputation sogar erhöhen (vgl. Bodewig, 1999, S.98).

3. Die Maximierung der Rücklaufquote als Ziel des Produktrückrufs

Die bisher geleisteten Ausführungen zu möglichen Konsequenzen eines Rückrufs implizieren, dass ein Verständnis zur Optimierung der Rückholung unsicherer Produkte vom Markt eine sinnvolle Investition für das Unternehmen darstellt. Kann ein Unternehmen nicht sicherstellen, dass alle unsicheren Produkte erfolgreich vom Markt entfernt werden, verbleiben viele dieser Produkte in den Händen der Verbraucher und können weiterhin eine Gefahr darstellen. Auch aufgrund des Interesses eines Unternehmens an einer langfristig gefestigten Marktposition sollte das oberste Ziel eines Produktrückrufs darin liegen, eine maximale Rückholung aller fehlerhafter Produkte aus dem Markt zu erreichen. (vgl. Van Venrooy, 2015, S.9)

3.1. Begriffserklärung Rücklaufquote

Das österreichische Bundesministerium für Arbeit, Soziales und Konsumentenschutz beschreibt die Rücklaufquote als den Anteil der zurückgeholten fehlerhaften Produkte im Verhältnis zu den ausgelieferten fehlerhaften Produkten (vgl. Bundesministerium für Arbeit, Soziales und Konsumentenschutz, 2017, S.2). Aufbauend auf dieser Definition kann festgehalten werden, dass die Rücklaufquote ein berechenbares Maß ist, dass eine konkrete Aussage über die Umsetzung des Rückrufes zu einem bestimmten Zeitpunkt darstellt (vgl. Van Venrooy, 2015, S. 60). Die Rücklaufquote ergibt sich aus dem Verhältnis zwischen der Anzahl der zurückgeholten Produkte aus dem Markt $P_{\text{zurück}}$ und der Gesamtheit der von einem Rückruf betroffenen Produkte P_{ges} . Damit wird die Rücklaufquote in folgender Weise berechenbar als:

$$\text{Rücklaufquote (\%)} = \frac{P_{\text{zurück}}}{P_{\text{ges}}} * 100 \tag{1}$$

Bezugnehmend auf die Forschungsfrage dieser Arbeit lässt sich die relevante Größe eindeutig definieren (vgl. ebd.). Die

Rücklaufquote kann sich grundsätzlich zwischen dem niedrigsten Wert 0 % und dem höchsten Wert 100 % bewegen (vgl. Van Venrooy, 2015, S.61).

Betrachtet man die maßgeblichen Richtlinien für einen Rückruf von Verbraucherprodukten in der deutschen Rechtsordnung, ist festzustellen, dass in ihnen keine eindeutige Definition einer Rücklaufquote zu finden ist. Im ProdSG findet sich lediglich die Forderung an Produktverantwortliche, eine von dem Verbraucherprodukt ausgehende Gefahr effektiv abzuwenden (vgl. §6 ProdSG). Einen Grund dafür, warum es keine gesetzliche Vorschrift über eine fest anzustrebende Rücklaufquote bei Verbraucherprodukten gibt, sieht der Produktrückruf-Experte Dr. Klindt und Wende (2014) darin, dass es dem Gesetzgeber vielmehr darauf ankommt, ob ein Unternehmen alle ihm möglichen und zumutbaren Kommunikationskanäle genutzt hat, um möglichst viele Endkunden zu erreichen (vgl. Klindt und Wende, 2014, S.89). Auch in den USA findet sich keine konkrete gesetzliche Aussage über eine anzustrebende Rücklaufquote bei Verbraucherprodukten.

3.2. Rücklaufquoten und ihre Bewertung

Eine Untersuchung der US-amerikanischen Verbraucherschutzbehörde CPSC ergab, dass die durchschnittliche Rücklaufquote bei Verbraucherprodukten in den USA im Jahr 2016 bei 33% lag (vgl. CPSC, 2016, S. 1). Allerdings ist in den USA ein Anstieg der Rücklaufquote im Zeitverlauf erkennbar. Im Jahr 1998 schwankte die durchschnittliche Rücklaufquote nach Schätzungen zufolge zwischen 10 und 30% (vgl. Coombs und Holladay, 2011, S. 412) während sie 1990 bei 2 bis 15% lag (vgl. Gibson, 1995, S.231). Auch bei Produktrückrufen in Deutschland sind nach Klindt und Wende (2014) mit niedrigen Rücklaufquoten von maximal 15 bis 20% zu rechnen (vgl. Klindt und Wende, 2014, S.2). Nach Angaben eines Behördenvertreters liegt die tatsächliche Rücklaufquote von Verbraucherprodukten in Deutschland allerdings deutlich geringer im einstelligen Bereich (vgl. Anhang B: Experteninterview mit Marc Göckeritz). Somit sind durchaus länderspezifische Unterschiede hinsichtlich der Rücklaufquote erkennbar.

Eine signifikante Abweichung von den allgemein niedrigen Rücklaufquoten bei Rückrufen von Verbraucherprodukten ist bei Rückrufen im Automobilbereich zu verzeichnen. Laut der US-Verkehrsbehörde NHTSA werden im Durchschnitt 75% aller von Rückrufen betroffenen Fahrzeuge erfolgreich vom Markt entfernt (vgl. NHTSA, 2016, S.1). Diese Tendenz lässt sich auch in der deutschen Automobilbranche erkennen, in der oft Rücklaufquoten von 70 bis 90% erreicht werden (vgl. Klindt und Wende, 2014, S.89). Diese hohen Rücklaufquoten in der Automobilindustrie resultieren aber allein aus der Besonderheit, dass Behörden einem Fahrzeughersteller in einem Rückruf die Anschriften aller Halter von kennzeichenpflichtigen Fahrzeugen bereitstellen können. Durch die sogenannte Halterabfrage können Automobilhersteller sicherstellen, dass sie alle betroffenen Fahrzeugbesitzer erreichen (vgl. Klindt und Wende, 2014, S.89, KBA, 2011, S.8). Gerade wegen dieser Besonderheit der behördlichen Bereitstellung der Halteranschriften können die hohen Rücklaufquoten in der Automobilindustrie nicht auf Verbraucherprodukte übertragen werden. Anhand der Abweichungen der Rücklaufquote zeigt sich hier jedoch bereits die Wichtigkeit, alternative Mittel zur Rückverfolgbarkeit der Produktbesitzer sicherzustellen, um ebenfalls alle betroffenen Konsumenten erreichen zu können.

Zu berücksichtigen ist bei der Bewertung niedriger Rücklaufquoten bei Verbraucherprodukten ebenfalls, dass die Berechnung zu einer systematischen Unterschätzung der Quote führen kann (vgl. Bodewig, 1999, S.91). Bei der Bewertung wird oft bei Rückrufen, die spät nach der Markteinführung der Produkte durchgeführt werden, nicht berücksichtigt, dass viele Produkte bereits nicht mehr im Gebrauch sind und keine Gefahr mehr darstellen. Die Rücklaufquote kann auch deshalb unterschätzt werden, weil einige Verbraucher defekte Produkte zum Teil selbst reparieren oder entfernen und aus diesem Grund nicht am Rückruf teilnehmen (vgl. ebd.).

Allerdings muss betont werden, dass selbst unter der Berücksichtigung einer möglichen systematischen Unterschätzung der Rücklaufquote die dargestellten Rücklaufquoten bei Verbraucherprodukten letztendlich enttäuschend gering sind, da es weiterhin zu einem Verbleiben von unsicheren Produkte kommt und die für Konsumenten bestehende Gefahr nicht vollkommen eliminiert wird (vgl. Bodewig, 1999, S.97).

3.3. Probleme bei der Erfüllung einer Maximierung der Rücklaufquote

Im Zuge der Maximierung der Rücklaufquote treten in der Regel komplexe Probleme auf die auf zwei Ebenen liegen (vgl. Van Venrooy, 2015, S. 71), Der Ebene des Rückrufmanagements und der Ebene des Verbraucherverhaltens.

Ein unzureichendes Produktrückrufmanagement

Ein frühzeitiges Wappnen eines Unternehmens für Rückrufsituationen ist eine notwendige und zum Teil gesetzlich vorgeschriebene Aufgabe der unternehmensinternen Organisation (vgl. Klindt und Wende, 2014, S.1). Konnte eine Gefahrenquelle nicht präventiv verhindert werden, muss ein Unternehmen durch eine zielgerichtete Organisation sicherstellen, dass es rechtzeitig und angemessen auf eine Rückrufsituation reagieren kann (vgl. Lenz, 2014, S.304). Hierbei können erhebliche Probleme aufkommen, beginnend bei Schwächen in der Identifikation von Fehlerquellen, einer mangelhaften Rückverfolgbarkeit von Produkten und Produktbesitzern, Probleme in der Entscheidung für Korrekturmaßnahmen bis hin zu Problemen in der Rückrufkommunikation an Konsumenten und Händler. Jegliche Fehlstrukturierung im Rückrufmanagement führt in Konsequenz zu unzureichenden Rücklaufquoten und folglich zu Kosten für das Unternehmen und Schäden für Konsumenten (vgl. Van Venrooy, 2015, S.71)). Somit zeigt sich, dass in Situationen des eigenen Produktversagens eine effiziente Organisationsfähigkeit eines Unternehmens unter Einhaltung der einschlägigen gesetzlichen Vorgaben ein unverzichtbares Grundgerüst zur Maximierung der Rücklaufquote darstellt.

Resistentes Verbraucherverhalten
Selbst für den Fall, dass die Vorbereitung und Durchführung vom Produktrückruf fehlerfrei organisiert wurde und alle betroffenen Konsumenten identifiziert und kontaktiert werden konnten, ist letztendlich das Verständnis und die Motivation der Konsumenten zur Teilnahme am Produktrückruf als ein entscheidender Faktor bei der Erfüllung einer maximalen Rücklaufquote anzuerkennen (vgl. Bodewig, 1999, S.94 ff.). Sogar bei Rückrufen in der Automobilindustrie, in denen durch die Halterabfrage ein Erreichen aller Automobilbesitzer sichergestellt werde kann, zeigt sich letzendes zwar eine hohe aber noch keine vollständige Rücklaufquote. Folglich ist die Rücklaufquotenentwicklung nicht nur als ein organisatorisches und rechtliches Problem einzustufen, sondern muss auch als verhaltenswissenschaftlich relevante Herausforderung angesehen werden, wobei allerdings hierzu bisher nur wenige Forschungsarbeiten vorliegen (vgl. CPSC, 2003, S.1). Ziel eines Rückrufs ist es möglichst viele Konsumenten zu erreichen und sie davon zu überzeugen am Rückruf teilzunehmen um somit eine maximale Schadensbegrenzung zu gewährleisten (vgl. Gröppel-Klein und Germelmann, 2009, S. 551). Ist die Rückrufkommunikation unzureichend oder unverständlich, wird sie von Verbrauchern möglicherweise ungehört oder ignoriert. Als unmittelbare Konsequenz ergibt sich eine ausbleibende Reaktion auf die Aufforderung zur Rückrufteilnahme (vgl. Van Venrooy, 2015, S.71). Welche Variablen einen Einfluss auf die Entscheidung der Verbraucher zur Teilnahme an Produktrückrufen ausüben, soll im Laufe der Arbeit näher untersucht werden.

Aus diesen Erkenntnissen lässt sich ableiten, dass für eine Maximierung der Rücklaufquote ein erfolgreiches Rückrufmanagement notwendig ist. Nur wenn es einem Unternehmen gelingt, alle betroffenen Konsumenten zu erreichen und sie zu einer Teilnahme an dem Produktrückruf zu motivieren, kann es eine maximale Rücklaufquote erreichen.

4. Die Regelung eines Rückrufmanagements in der deutschen und US-amerikanischen Rechtsordnung

Wie bereits im Kapitel 3.2 erwähnt, sind durchaus Unterschiede in den erzielten Rücklaufquoten bei Rückrufen von Verbraucherprodukten zwischen Deutschland und den USA erkennbar. Aufgrund der Erkenntnis, dass ein erfolgreich durchgeführtes Rückrufmanagement einen Einfluss auf die Rücklaufquotenentwicklung hat, kann die länderspezifische Regelung eines Rückrufmanagements auch einen Einfluss auf den Rückruferfolg ausüben. Aus diesem Grund soll im folgenden Kapitel die Regelung eines Rückrufmanagements in der deutschen und US-amerikanischen Rechtsordnung untersucht werden.

Die Notwendigkeit eines Rückrufmanagements
Das Rückrufmanagement umfasst die direkte Durchführung eines Rückrufs und alle Vorbereitungen, die zur erfolgreichen Umsetzung präventiv getroffen werden müssen (vgl. Klindt und Wende, 2014, S.3). In Deutschland wird der Aufbau im ProdSG gesetzlich vorgeschrieben. Unternehmen müssen „jeweils im Rahmen der Geschäftstätigkeit Vorkehrungen für geeignete Maßnahmen zur Vermeidung von Risiken zu treffen“ (§ 6 Abs.2 ProdSG). Der deutsche Gesetzgeber verpflichtet Unternehmen zum Aufbau präventiver Organisationsstrukturen, um ein zielgerichtetes Reagieren bei Produktunsicherheiten zu ermöglichen. Hierbei stellt der Rückruf die Ultima Ratio neben einer Vielzahl anderer Möglichkeiten dar (vgl. Wilrich, 2012, S. 173 ff.). Hingegen beruht die Notwendigkeit eines Rückrufmanagements in Amerika nicht wie in Deutschland unmittelbar auf einem direkten öffentlich-rechtlichen Handlungsauftrag, sondern lässt sich mittelbar aus produkthaftungs-rechtlichen Aspekten ableiten. Der Antrieb zum Aufbau eines Rückrufmanagements erfolgt hier als konsequente Vermeidung von exzessiven Haftungsansprüchen im Falle von eintretenden Verletzungen (vgl. Haberland, 2016, S.188ff.).

4.1. Die Regelung eines Rückrufmanagements in der deutschen Rechtsordnung

Die Rückrufregelung unsicherer Verbraucherprodukten basiert in der deutschen Rechtsordung auf dem ProdSG als nationale Umsetzung der europäischen Richtlinie 2001/95/EG (vgl. Rödl & Partner, 2014, S.68). In Deutschland konzipieren die Bundesländer jeweils eine eigenständige Marktüberwachung auf Grundlage der gesetzlichen Vorgaben. Je nach Bundesland sind verschiedene Behörden für die Überwachung von Rückrufen verantwortlich (BMUB, 2016, S.1). Die Abbildung 1 zeigt den grundsätzlichen Ablauf eines Rückrufmanagements.

Produktbeobachtung & Ergreifen vorbeugender Maßnahmen
Das ProdSG fordert von Unternehmen im Rahmen ihrer Produktbeobachtungspflicht Beschwerden von Verbrauchern zu prüfen, das Führen eines Beschwerdebuchs und die Durchführung gebotener Stichproben (vgl. ProdSG § 6 Abs.3). Um Informationen über Produktfehler zu bekommen, stehen Unternehmen verschiedene Quellen zur Verfügung. Unternehmen können durch Verbraucher und Händler über unsichere Produkte informiert werden oder Produktfehler selbst in Eigenkontrollen entdecken (vgl. Foodwatch, 2017, S. 47). Auch Behörden können im Rahmen der Marktüberwachung Unternehmen über eine sicherheitskritische Situation kontaktieren. Es liegt im Ermessen der jeweiligen Behörde, ob sie einen Rückruf, eine Warnung oder andere Maßnahmen anordnen (§ 26 Abs. 2 S.1,2 Nr.7 Alt.2 ProdSG). Entdeckt eine Behörde hingegen, bevor ein Unternehmen selbst darauf reagiert, dass von einem Produkt eine ernste Gefahr für Leib und Leben ausgeht, hat sie die gesetzliche Pflicht zum Einschreiten: Eine sogenannte Ermessensreduzierung auf Null (vgl. § 26 Abs.4 ProdSG). Die Behörde wird in diesem Fall einen Rückruf anordnen. Darüber hinaus sind Unternehmen gesetzlich verpflichtet, Vorkehrungen zu treffen, um einen Rückruf erforderlichenfalls erfolgreich durchführen zu können (§6 Abs.2 ProdSG). Um in einem Rückruf alle betroffenen Produkte und deren Besitzer identifizieren zu können, müssen im Vorfeld Maßnahmen getroffen werden, die eine Rückverfolgbarkeit sicherstellen. Die Gesetzesbegründung

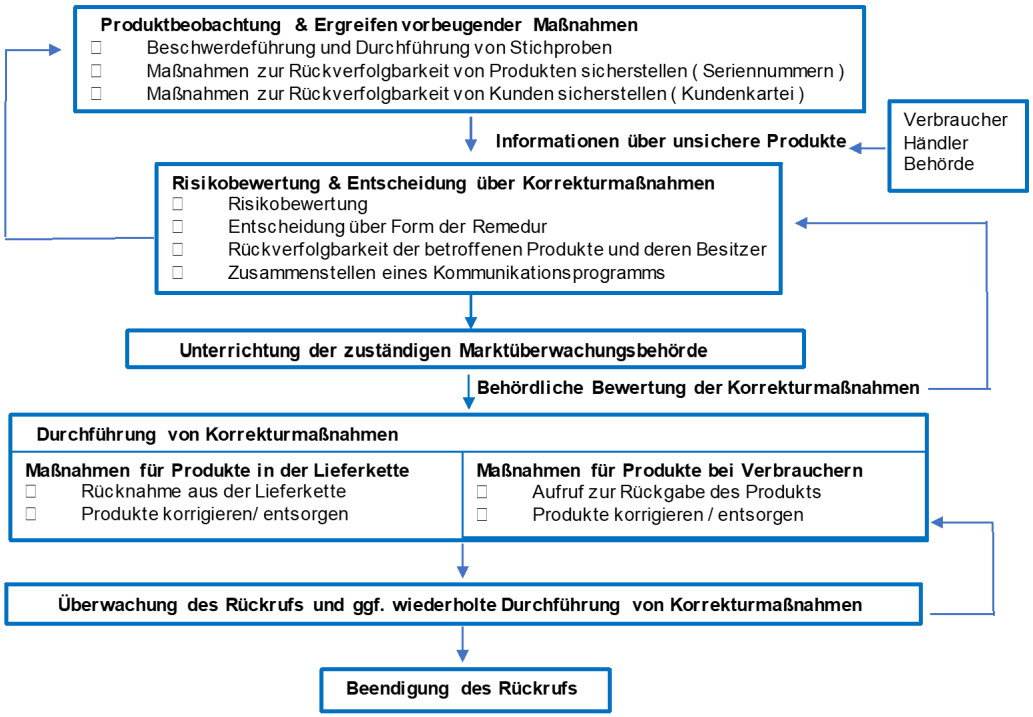


Abbildung 1: Rückrufmanagement in der deutschen Rechtsordnung (eigene Darstellung, in Anlehnung an: ProSafe, 2011, S.13)

schlägt die Vergabe von Seriennummern und die Führung von Kundendateien vor (vgl. BT-Drucksache 15/1629, 2003, S.29). Wie Unternehmen die Rückverfolgbarkeit letztendlich sicherstellen, liegt im Ermessen der Unternehmen (vgl. Klindt und Wende, 2014, S.87).

Risikobewertung

Hat ein Unternehmen festgestellt, dass ein Produkt eine Gefahr für Verbraucher darstellen könnte oder wurde es von einer Quelle auf eine potenzielle sicherheitsrelevante Produktauffälligkeit aufmerksam gemacht, besteht der nächste Schritt darin, die vom Produkt ausgehende Gefahr zu bewerten. Wie Unternehmen Produktrisiken bewerten, wird nicht gesetzlich vorgeschrieben. Jedoch wird im ProdSG die Risikobewertung nach dem RAPEX Leitfaden der Europäischen Kommission als verbindliche Vorgabe für Behörden festgelegt (vgl. §26 Abs.4 ProdSG). Ableitend wird auch Unternehmen empfohlen den RAPEX Leitfaden der eigenen Risikobewertung zugrunde zu legen (vgl. Klindt und Wende, 2014, S.64). Hier wird der Risikograd eines Produkts aus der Kombination aus dem Schweregrad der möglichen Verletzung und der Eintrittswahrscheinlichkeit einer Schädigung der Konsumenten während der voraussichtlichen Lebensdauer eines Produktes ermittelt (vgl. BAuA, 2017b).

Ein ernstes Risiko liegt dann vor, wenn ein Produkt Verletzungen hervorrufen kann, die stationäre Behandlungen erfordern, zu schweren Verletzungen oder zum Tod führen können und eine hohe Eintrittswahrscheinlichkeit haben (ebd.).

Entscheidung über Korrekturmaßnahmen

Auf Basis der Risikoeinschätzung wird entschieden, welche Korrekturmaßnahmen getroffen werden müssen, um Risiken für Verbraucher zu vermeiden. Im Falle eines ernsten Risikos fordert das ProdSG Unternehmen auf, angemessene Maßnahmen zu treffen, die bis zur Rücknahme, Warnung oder zum Rückruf reichen (vgl. ProdSG §6 Abs.2). Wenn die betroffenen Produkte die Vertriebskette überwiegend schon verlassen haben und in den Händen der Verbraucher vermutet werden, ist ein Rückruf unumgänglich (vgl. Van Venrooy, 2015, S.81). Gleichzeitig muss ein Unternehmen sicherstellen, dass unsichere Produkte, die sich noch in der Vertriebskette befinden, ebenfalls erfolgreich entfernt werden (vgl. ProSafe, 2011, S.12).

Entscheidet sich ein Unternehmen für einen Rückruf, liegt eine Entscheidung in der Wahl für angebotene Kompensation. Unternehmen können Konsumenten eine Reparatur, einen Umtausch, eine teilweise oder vollständige Kaufpreisrückerstattung, Gutscheine oder Ermäßigungen anbieten (vgl. Liu et al., 2016, S.79). Welche Kompensationsart bei einem Rückruf gewählt wird, liegt im Ermessen des jeweiligen Unternehmens selbst, da es hierzu keine gesetzlichen Vorgaben zu beachten hat. Häufig stellt sich in diesem Zusammenhang die Frage der Kostentragung für angebotene Reparaturmaßnahmen. In der deutschen Rechtsordnung besteht keine gesetzliche Kostentragungspflicht für Unternehmen bei Reparaturmaßnahmen im Rahmen einer Rückrufaktion (vgl. Haberland, 2016, S. 60).

Welche Kommunikationsmaßnahmen ein Unternehmen ergreift, um alle betroffenen Konsumenten zu erreichen, hängt davon ab, wie effektiv ein Unternehmen präventive

Maßnahmen zur Sicherstellung der Rückverfolgbarkeit ergriffen hat. Falls ein Unternehmen gewährleisten kann, dass ihnen alle Kontaktdaten der betroffenen Konsumenten zur Verfügung stehen, können sie die betroffenen Produktnutzer direkt kontaktieren (vgl. Klindt und Wende, 2014, S.86). Ist eine individuelle Kommunikation nicht möglich wird ein öffentlicher Rückruf unumgänglich. Hierbei entscheidet das Unternehmen zunächst selbst, welche medialen Kanäle es nutzt, um öffentlich vor den betroffenen Produkten zu warnen. Es gibt keine gesetzlichen Vorgaben, welche Kommunikationskanäle für einen Rückruf genutzt werden müssen um zu gewährleisten, dass alle betroffenen Verbraucher informiert werden. Ob ein Unternehmen Kanäle in den sozialen Medien nutzt oder Aushänge im Einzelhandel veranlasst– bei diesen Fragen verfügen Unternehmen über große Ermessensspielräume (vgl. Foodwatch, 2017, S.21).

Unterrichtung der zuständigen Marktüberwachungsbehörden

Nach §6 Abs.4 ProdSG sind Hersteller von Verbraucherprodukten verpflichtet, die zuständige Behörden zu unterrichten, wenn sie wissen, dass ein Produkt ein Risiko für die Gesundheit und Sicherheit von Personen darstellt. Im Rahmen seiner Meldepflicht muss ein Unternehmen Informationen über das betroffene Produkt, den Risikograd, bereits getroffene oder geplante Korrekturmaßnahmen und bekannt gewordene Vorfälle angeben (vgl.§6 Abs.4 ProdSG). Die zuständige Behörde muss wiederum die Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA) über jeden eingehenden Rückruf informieren (§6 Abs.4 S.2 ProdSG). Die BAuA agiert als nationaler Kommunikationsknotenpunkt und veröffentlicht die ihr bekannt gewordenen Produktrückrufe auf ihrer Rückrufplattform (vgl. BAuA, 2017c).

Im Anschluss prüft die zuständige Behörde, ob sie die vom Unternehmen getroffenen Korrekturmaßnahmen als ausreichend empfindet und kann unter Umständen weitere Maßnahme anfordern (vgl. Anhang B). Die Organisation Foodwatch stellt in einer aktuellen Studie fest, dass sich deutschen Behörden bei der Überprüfung „meist mit dem Nachweis zufriedengeben, dass eine Pressemitteilung an die Deutsche Presseagentur verschickt wurde - in viele weitere Details aber mischen sie sich nicht ein“ (Foodwatch, 2017, S.21). Trotz der Tatsache, dass sich Foodwatch in ihrer Studie auf Lebensmittel bezieht, können die Aussagen über das Behördenverhalten auch auf die Rückrufpraxis bei Verbraucherprodukten bezogen werden, da in vielen Bundesländern die zuständigen Behörden für die Marktüberwachung von Lebensmitteln und auch Verbraucherprodukten zuständig sind.

Durchführung der Korrekturmaßnahmen, Überwachung und Beendigung des Rückrufs

Nachdem die vom Unternehmen getroffenen Korrekturmaßnahmen von der Behörde geprüft und genehmigt oder gegebenenfalls durch eine behördliche Nachforderung geändert wurden, kommt es zur Umsetzung der Maßnahmen. Hauptverantwortlicher für die Durchführung eines Rückrufs ist das Unternehmen (vgl. ProSafe, 2011, S.20). Zum einen muss es sicherstellen, dass der Teil der betroffenen Produkte, der sich teilweise noch in der Lieferkette befindet, erfolgreich entfernt wird. Hierbei sind Händler verpflichtet sicherzustellen, dass keine der vom Rückruf betroffenen Produkte mehr im Handel verkauft werden (vgl. §6 Abs.5 S.1 ProdSG). Je nachdem ob ein Unternehmen einen stillen oder einen öffentlichen Rückruf durchführt, muss es die Rückrufaufforderung individuell oder über die öffentlichen Medien an betroffene Konsumenten verschicken. Die Überwachung, ob die Durchführung der getroffenen Korrekturmaßnahmen zu einer positiven Rücklaufquotenentwicklung führt, liegt in den Händen der Unternehmen. Nach Aussagen eines Vertreters des MULNV NRW wird die Entwicklung der Rücklaufquote von der zuständigen Behörde nicht kontrolliert. Auch besteht keine Verpflichtung, dass ein Unternehmen die Entwicklung der Rücklaufquote dokumentieren muss (vgl. Anhang B). Die Entscheidung, ob im Falle einer niedrigen Rücklaufquote, ein Aufruf zur Rückgabe ein zweites Mal durchgeführt wird, liegt im alleinigen Ermessen des Unternehmens. In der Regel ordnet die zuständige Behörde keine zweite Durchführung von Korrekturmaßnahmen an. Auch die Entscheidung, wann ein Rückruf beendet wird, entscheidet das Unternehmen selbst (vgl. Anhang B).

Die Regelung eines Rückrufmanagements in der US-amerikanischen Rechtsordnung

Die Regulierung eines Rückrufs von Verbraucherprodukten basiert im US-Recht auf dem Gesetz zur Sicherheit von Konsumgütern CPSIA (Consumer Product Improvement Act), das als umfassende Überarbeitung des CPSA (Consumer Product Safety Act) im Jahr 2008 in Kraft trat. Die CPSC wacht als unabhängige bundesstaatliche Verbraucherschutzbehörde über die Einhaltung der gesetzlichen Bestimmungen und stellt die bundesweit zuständige Behörde für meldepflichtige Rückrufe dar (vgl. Haberland, 2016 , S.148). Grafisch lässt sich der Ablauf eines Rückrufmanagements in der US-amerikanischen Rechtsordnung wie folgt darstellen (Abbildung 2).

Produktbeobachtung (continuing duty to test) und Ergreifen vorbeugender Maßnahmen

Die fortlaufende Produktbeobachtung (continuing duty to test) und das Ergreifen vorbeugender Maßnahmen stellt im US-amerikanischen Recht einen entscheidenden Bestandteil der betrieblichen Organisation dar. Zwar besteht in den USA keine unmittelbare Produktbeobachtungspflicht, jedoch haben Hersteller von Verbraucherprodukten eine allgemeine Pflicht zur Warnung (post sale duty to warn) sowie eine Meldepflicht bei der CPSC, was faktisch eine kontinuierliche Produktbeobachtungspflicht voraussetzt (vgl. Haberland, 2016, S.188). Denn durch eine Verletzung der Warn- und Meldepflicht, drohen einem Unternehmen exzessive Bußgeldvollstreckungen und Strafschadensersatzansprüche (punitive damages) (vgl. 15U.S.C. §2069, §2070). Zur Vermeidung dieser ist die Produktbeobachtung eine essentielle Voraussetzung. Für die Produktbeobachtung und der Sicherstellung der Rückverfolgbarkeit von Produkten und deren Besitzern schlägt die CPSC Maßnahmen, wie die Prüfung von Verbraucherbeschwerden, Durchführung von Qualitätskontrollen, den Gebrauch von Produktregistrierungskarten oder die Vergabe von Seriennummern vor. Die genaue Ausgestal-

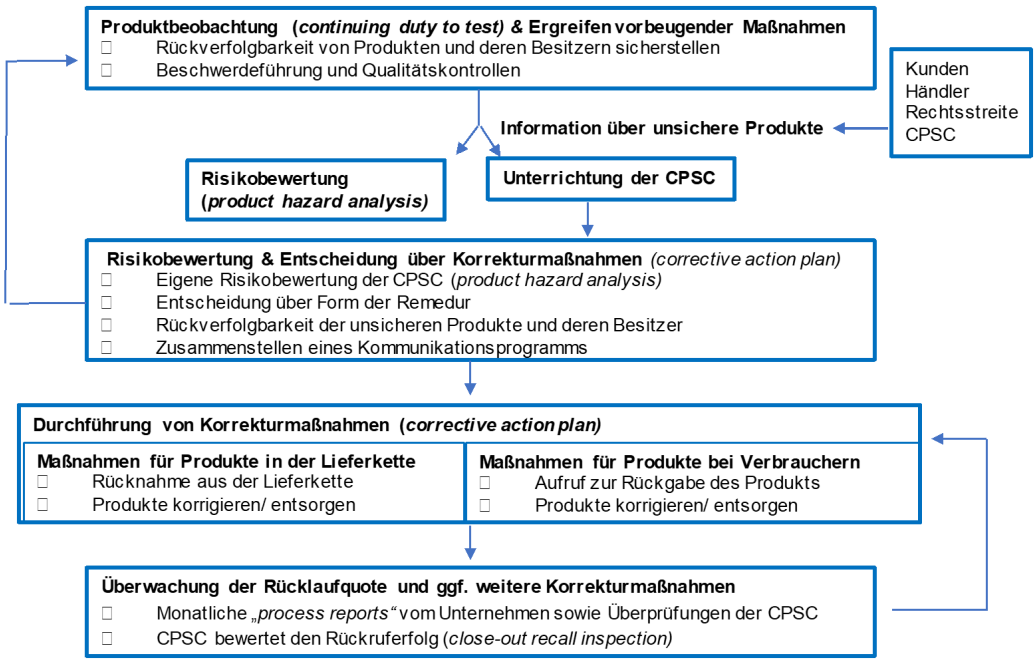


Abbildung 2: Rückrufmanagement in der US-amerikanischen Rechtsordnung (eigene Darstellung)

tung liegt dabei im Ermessen des jeweiligen Unternehmens (CPSC, 2012, S. 28). Ein Unternehmen kann von Kunden, Mitgliedern aus der Wertschöpfungskette oder von der CPSC über Produktunsicherheiten aufmerksam gemacht werden. Stellt die CPSC fest, dass ein Produkt gegen gesetzliche Produktstandards verstößt, teilt sie dies dem Unternehmen in einem Letter of Advice (LOA) mit. Ein LOA enthält die Aufforderung, konkrete, von der CPSC als notwendig erachtete Maßnahmen (wie einen Rückruf) zu initiieren. Bestätigt das Unternehmen die Verletzung der Produktstandards, erarbeitet es zusammen mit der CPSC einen konkreten Rückrufplan (correcitve action plan) (vgl. CPSC, 2012, S.18)

Risikobewertung und Unterrichtung der Verbraucherschutzbehörde CPSC

In den USA besteht eine strenge Meldepflicht an die CPSC bei unsicheren Verbraucherprodukten, welche auf zwei möglichen Bedingungen beruht. Nach Section 37 des CPSA ist eine Meldung bei der CPSC verpflichtend, sobald drei oder mehr Haftungsansprüche, die einen Fehler an einem bestimmten Produkt betreffen, innerhalb einer Zweijahresfrist abgeschlossen wurden (vgl. 15 U.S.C. §2084(a)). Die überwiegende Anzahl der Meldungen an die CPSC erfolgt jedoch aufgrund der Anforderungen nach Section 15(b) des CPSA. Unternehmen sind hiernach verpflichtet, die CPSC umgehend zu benachrichtigen, sobald sie Informationen darüber erhalten, die berechtigten Grund zur Annahme geben, dass ein Produkt bestimmten Standards nicht entspricht oder einen Fehler aufweist, von dem eine erhebliche Gefahr (substantial product hazard) ausgeht oder eine unangemessene Gefahr ernsthafter Verletzungen oder Tod erzeugt (vgl. 15 U.S.C. §2064(b)).

Somit muss das meldepflichtige Unternehmen im Vorfeld bewerten, ob ein potenzieller Produktfehler ein substantial product hazard darstellen kann. Dieser lässt sich nach Section 15(a) (2) des CPSA nach der Schwere der Gefahr, der Wahrscheinlichkeit einer Verletzung und der Anzahl fehlerhafter Produkte im Handel bestimmen. Generell besteht ein substantial product hazard, wenn die potenzielle Gefährdung schwerwiegend oder sehr wahrscheinlich ist (vgl. 15 U.S.C. §2064). Sobald die Entscheidung zu der Risikoeinschätzung vorliegt ist das Unternehmen verpflichtet die CPSC innerhalb von 24 Stunden zu informieren (vgl. CPSC, 2013a, S.43). Stellt die CPSC fest, dass ein Unternehmen die Meldeanforderungen nach dem CPSA nicht richtig oder nicht rechtzeitig eingehalten hat, können Unternehmen mit exzessiven Sanktionen belegt werden, welche 2008 durch den CPSIA noch einmal deutlich verschärft wurden. So können Verstöße gegen die Meldepflicht mit empfindlichen Geldbußen von bis zu 15 Millionen Dollar und strafrechtlichen Verfolgungen geahndet werden (vgl. CPSC, 2013a, S.11). Um die enorm hohen Sanktionen zu vermeiden rät die CPSC Unternehmen dazu, bereits während der eigenen Risikobewertung die Behörde über den Verdacht eines gefährlichen Produktfehlers zu informieren (CPSC, 2012, S.1)

Gemeinsame Entscheidung über Korrekturmaßnahmen (corrective action plan)

Im Anschluss nimmt die CPSC eine eigene Risikobewertung (product hazard analysis) des gemeldeten Produktes vor, die nach denselben Kriterien der Section 15 (a) (2) des CPSA erfolgt, welche ebenfalls bei der unternehmensinternen Risikobewertung herangezogen werden (vgl. 15 U.S.C. §2064). Stellt die CPSC fest, dass von einem Produktfehler ein sub-

stantial product hazard ausgeht, wird hierbei zwischen drei Schweregraden unterschieden. Je nach Schweregrad des Produktfehlers gibt die CPSC unterschiedliche Vorgaben über die Art der Maßnahmen und zu nutzenden Informationskanäle (vgl. CPSC, 2012, S.14). Beispielsweise liegt ein Class A Hazard vor, wenn das Produkt unter einer sehr hohen Wahrscheinlichkeit schwere Verletzungen oder den Tod verursachen kann. Dieser Schweregrad erfordert sofortige und umfassende Korrekturmaßnahmen (vgl. CPSC, 2012, S.14). Entscheiden sich das Unternehmen und die CPSC zur Durchführung eines Rückrufs, wird im nächsten Schritt gemeinsam ein konkreter Rückrufplan (corrective action plan) erarbeitet. Je nach dem klassifizierten Schweregrad wird gemeinsam über die Form der Kompensation entschieden und ein Plan über die Rückrufkommunikation ausgearbeitet (vgl. CPSC, 2012, S.17). Welche Kompensationsart bei einem Rückruf gewählt wird, entscheidet das Unternehmen in Abstimmung mit der CPSC. Sieht die Behörde Reparaturmaßnahmen der defekten Produkte oder das Ausstellen von Gutscheinen als Anreiz für die Rückgabe betroffener Produkte als notwendig an, haben Unternehmen die Pflicht alle anfallenden Kosten zu übernehmen. Die Rechtslage in den USA verlangt eine vollständige Kostentragung durch das Unternehmen für etwaige Kompensationen im Rahmen eines Rückrufs (vgl. Haberland, 2016, S.196).

Welche Kommunikationsmaßnahmen ein Unternehmen ergreift um alle betroffenen Konsumenten zu erreichen, hängt davon ab, wie erfolgreich die Rückverfolgbarkeit der Produkte und deren Besitzer im Vorfeld sichergestellt wurde. Sind nicht alle Kontaktdaten der betroffenen Konsumenten verfügbar, ist eine öffentliche Kommunikation des Rückrufs notwendig (vgl. CPSC, 2012, S.20). In diesem Fall veröffentlichen die CPSC und das Unternehmen gemeinschaftlich eine Pressemitteilung. Hierbei gibt die CPSC Vorgaben über die Formulierung und des Inhalts der Pressemitteilung, welche ebenfalls als Grundlage für die Ausgestaltung aller weiteren Bekanntmachungen herangezogen werden muss (vgl. 15 U.S.C. §2064 (i) (2)). Die Entscheidung welche medialen Kanäle ein Unternehmen für die Produktrückrufkommunikation nutzt, wird gemeinschaftlich mit der CPSC getroffen (vgl. CPSC, 2012, S.18). Das entscheidende Kriterium für die Wahl der Informationskanäle ist der zuvor bewertete Schweregrad des betroffenen Produkts. Je höher der Schweregrad, desto mehr Vorgaben sind hinsichtlich der zu nutzenden Kommunikationskanäle zu befolgen. So fordert die CPSC im Falle eines Class A Hazard eine maximale Nutzung von Informationskanälen zur Bekanntgabe des Rückrufs (vgl. CPSC, 2012, S.14). Jede öffentliche Mitteilung muss im Vorfeld von der CPSC überprüft und genehmigt werden (vgl. ebd.). Somit zeigt sich, dass Unternehmen ein geringer Ermessensspielraum bei der Ausgestaltung der Korrekturmaßnahmen eingeräumt wird, da jede Entscheidung im Produktrückruf einvernehmlich mit der CPSC getroffen werden muss.

Durchführung der Korrekturmaßnahmen, Überwachung und Beendigung des Rückrufs

Das Unternehmen muss einerseits erreichen, dass alle betroffene Kunden am Rückruf teilnehmen und dass alle unsicheren Produkte, die sich noch in der Lieferkette befinden, erfolgreich entfernt werden. Die Überwachung der Durchführung von Korrekturmaßnahmen in der Lieferkette ist sehr wichtig, da der Verkauf, Import oder die Distribution von Produkten, die von einem Produktrückruf betroffen sind, gemäß des CPSIA für jeden Händler, Hersteller oder Importeur verboten ist. Jeder Wirtschaftsakteur kann bei Nichteinhaltung des Gesetzes mit erheblichen strafrechtlichen Sanktionen belegt werden (vgl. 15 U.S.C. §2068). Die Durchführung des Rückrufs wird vom betroffenen Unternehmen und von der CPSC überwacht. Die CPSC führt regelmäßige Überprüfungen und Kontrollbesuche im Handel durch, um zu überprüfen, ob der corrective action plan vom Unternehmen angemessen umgesetzt wird und ob keine unsicheren Produkte mehr verkauft werden (vgl. CPSC, 2012, S.25f.). Das Unternehmen ist verpflichtet der CPSC einen monatlichen Bericht (monthly progress reports) vorzulegen, auf dessen Grundlage der Erfolg des Rückrufs von der CPSC bewerten wird. Hierbei müssen Unternehmen die CPSC über die Anzahl der erfolgreich vom Markt entfernten Produkte und bekannt gewordene Vorfälle während des Rückrufs informieren (vgl. ebd.) Wenn ein Unternehmen zu dem Entschluss kommt, dass es so viele unsichere Produkte, wie es in seiner Macht stand, vom Markt entfernt hat, kann es der CPSC einen finalen Bericht vorlegen, indem es beantragt, dass die Kommission den Rückruf als beendet erklärt. Erst nach einer abschließenden Bewertung der CPSC über den Erfolg der Korrekturmaßnahmen (close-out recall inspection) wird der Rückruf offiziell beendet. Falls die CPSC der Meinung ist, dass der erarbeitete correcitve action plan nicht die erwünschte Wirkung erzielt, kann es die Durchführung von umfassenderen Korrekturmaßnahmen anordnen (vgl. CPSC, 2012, S.26).

5. Gegenüberstellung der Regelung eines Rückrufmanagements in der deutschen und US-amerikanischen Rechtsordnung

Nachdem nun die jeweiligen Vorgänge eines Rückrufmanagements in den beiden Rechtsordnungen einzeln vorgestellt wurden, soll sich der folgende Teil der Arbeit mit dem Vergleich der deutschen und US-amerikanischen Rechtsordnung zur Regelung eines Rückrufmanagements befassen. Es dient dazu, sowohl Gemeinsamkeiten als auch entscheidende Unterschiede der Regelung in beiden Rechtssysteme darzustellen. Dabei werden wesentliche Schritte im Rückrufmanagement der Länder aufgegriffen, näher betrachtet und gegenübergestellt. In Tabelle 1 soll die Regulierung der wesentlichen Schritte im Rückrufmanagement in beiden Rechtssystemen kurz gegenübergestellt werden.

5.1. Gemeinsamkeiten

Die Untersuchung der Regelung eines Rückrufmanagements in den USA und Deutschland hat ergeben, dass durchaus Gemeinsamkeiten in einigen Schritten des Rückrufmanagements zwischen den beiden Rechtsordnungen zu erkennen sind.

Tabelle 1: Gegenüberstellung der Regelung eines Rückrufmanagements in der deutschen und US-amerikanischen Rechtsordnung (Quelle: eigene Darstellung)

Schritt im Rückrufmanagement	Deutschland	USA	Unterschied
Produktbeobachtung	Produktbeobachtungspflicht		nein
Ergreifen vorbeugender Maßnahmen	vergleichbare Vorschläge zur Sicherstellung der Rückverfolgbarkeit von Produkten und deren Besitzern		nein
Risikobewertung	vergleichbare Bewertungskriterien		nein
Unterrichtung der Behörde	Meldepflicht: ja Zeitpunkt: nach Entscheidung über Korrekturmaßnahmen Sanktionen: moderat	Meldepflicht: ja Zeitpunkt: vor Entscheidung über Korrekturmaßnahmen Sanktionen: sehr hoch	ja
Entscheidung über Korrekturmaßnahmen	Entscheidungsträger: Unternehmen und Beurteilung durch Behörde Zeitpunkt: vor der Unterrichtung der zuständigen Behörde	Entscheidungsträger: CPSC und Unternehmen Zeitpunkt: nach der Unterrichtung der Behörde	ja
Entscheidung über Remedur	Entscheidungsträger: Unternehmen Kostentragungspflicht: nein	Entscheidungsträger: CPSC und Unternehmen Kostentragungspflicht: ja	ja
Entscheidung über Kommunikationsmaßnahmen	Entscheidungsträger: Unternehmen und Beurteilung durch Behörde Vorgaben: nicht einheitlich Kontrolle d. Behörde: nicht einheitlich	Entscheidungsträger: CPSC und Unternehmen Vorgaben: umfangreich Kontrolle d. Behörde: streng	ja
Gestaltung der Rückrufanzeige	Vorgaben: nicht einheitlich Kontrolle d. Behörde: nicht einheitlich	Vorgaben: umfangreich Kontrolle d. Behörde: streng	ja
Durchführung & Überwachung des Rückrufs	Überwachung: Unternehmen Kontrolle d. Behörde: keine	Überwachung: Unternehmen und CPSC Kontrolle d. Behörde: streng	ja
Beendigung des Rückrufs	Entscheidungsträger: Unternehmen	Entscheidungsträger: CPSC	ja

5.1.1. Produktbeobachtung und Ergreifen vorbeugender Maßnahmen

Sowohl in Deutschland als auch in den USA stellen die Produktbeobachtung und das Ergreifen vorbeugender Maßnahmen organisatorische Maßnahmen dar, welche präventiv geschaffen werden müssen, um eine erfolgreiche Umsetzung des Rückrufs und somit einer maximalen Rücklaufquote gewährleisten zu können.

Beide Rechtsansichten sind sich darüber einig, dass die Produktbeobachtung (continuing duty to test) einen entscheidenden Bestandteil der betrieblichen Organisation darstellt. Der einzige Unterschied besteht darin, dass sich die Produktbeobachtungspflicht in den USA nicht wie in Deutschland unmittelbar aus einem Gesetz ergibt, sondern sich mittelbar aus produkthaftungsrechtlichen Aspekten ableitet (vgl. Haberland, 2016, S.188).

Durch die Produktbeobachtung verfolgt ein Unternehmen aktiv das Verhalten seines Produkts im Markt, um auf Produktfehler frühestmöglich korrigierend Einfluss nehmen zu können. In Bezug auf den Rückruf stellt eine Produktbeobachtung somit eine notwendige Aufgabe dar, um sicherheitsrelevante Fehler an Produkten überhaupt erst zu erkennen, die einen Produktrückruf auslösen können. (vgl. Klindt und Wende, 2014, S.64). Um im Falle eines Rückrufs unsichere Produkte identifizieren zu können, ist es notwendig im Vorfeld präventive Maßnahmen zu ergreifen, die eine hohe Rückverfolgbarkeit der Produkte sicherstellen. Denn je präziser zugeordnet werden kann, in welchen ausgelieferten Produkten das jeweilige Sicherheitsproblem liegt und wo sich die unsicheren Produkte befinden, umso präziser und eingegrenzter kann der Produktrückruf durchgeführt werden. Nun kann angenommen werden, dass sich die Rückverfolgbarkeit von Verbraucherprodukten als Herausforderung erweist, wenn sie in großen Stückzahlen in den globalen Markt gebracht werden. Unternehmen können über den Aufenthalt der Produkte innerhalb der Lieferkette noch relativ zuverlässig Aussagen treffen. Gelangen Produkte jedoch an den Endkunden, wird es schwer die Produktdaten mit den jeweiligen Kundendaten zu erfassen (vgl. Klindt und Wende, 2014, S.87).

Insgesamt ergeben sich für Unternehmen in beiden Rechtsordnungen ähnlich große Freiheitsgrade in der Entscheidung, wie sie die Rückverfolgbarkeit von Produkten und Kunden sicherstellen. Es werden vergleichbare Vorschläge zur Rückverfolgbarkeit gemacht, wie die Vergabe von Seriennummern und das Führen einer Kundendatei (vgl. BT-Drucksache 15/1629, 2003, S.29, CPSC, 2012, S. 28). Nur bei Produkten für Kleinkinder und Säuglingen bestehen in den USA strenge gesetzliche Anforderungen für die Kundenregistrierung zur Sicherstellung der Rückverfolgbarkeit von Produktbesitzern, wie die Pflicht zur Führung von Produktregistrierungskarten (CPSC, 2013b, S.1).

5.1.2. Risikobewertung

Wurde ein Unternehmen auf eine sicherheitsrelevante Produktauffälligkeit aufmerksam gemacht oder hat das Unternehmen einen Produktfehler selbst im Rahmen der Produktbeobachtung identifiziert, muss das potenzielle Risiko des betroffenen Produkts systematisch bestimmt werden (vgl. ProSafe, 2011, S.19). In den Rechtsordnungen sind Übereinstimmungen über die Vorgehensweise der Risikobeurteilung erkennbar. So wird sowohl in den USA als auch in Deutschland vorgeschrieben, dass der Risikograd des betroffenen Produkts aus der Schwere der Gefahr unter Berücksichtigung der Eintrittswahrscheinlichkeit der potenziellen Gefahr ermittelt wird (vgl. 15 U.S.C. §2064, BAuA, 2017b). In beiden Rechtsordnungen liegt die Risikobewertung zuerst im Verantwortungsbereich des Unternehmens, während im späteren Verlauf von der CPSC und von deutschen Behörden eine eigenständige Risikobewertung nach derselben Vorgehensweise durchgeführt wird (vgl. 15 U.S.C. §2064, §26 Abs.4 ProdSG). Der ermittelte Risikograd ist maßgebend für den weiteren Umgang mit dem unsicheren Produkt. Er ist letztlich ein wichtiger Indikator, um zu bestimmten, ob ein Rückruf gestartet werden muss und welche Korrekturmaßnahmen erforderlich sind. Ferner hängt von der Risikobewertung ab, ob Behörden an einer Kontrolle des Rückrufs interessiert sind. In beiden Rechtsordnungen beschränkt sich der behördliche Fokus auf Rückrufe von gefährlichen oder nicht konformen Produkten. Rückrufe als Maßnahme zur Sicherstellung unternehmensinterner Qualitätsstandards werden von Behörden in der Regel nicht unterstützt (vgl. Van Venrooy, 2015, S.92).

5.2. Unterschiede

Allerdings sind auch evidente Unterschiede in der Regelung des Rückrufmanagements zwischen der Rechtsordnung in Deutschland und den USA sichtbar. Die Unterschiede zeigen sich in einer Veränderung im Ablauf der Prozesskette als auch in einer unterschiedlichen Stärke der Regulierungsdichte in einigen Schritten des Rückrufmanagements.

5.2.1. Unterrichtung der zuständigen Behörde

Zwar besteht allgemein in beiden Rechtsordnungen eine Meldepflicht, wenn Unternehmen Kenntnis über unsichere Produkte haben, dennoch lassen sich Unterschiede hinsichtlich des Zeitpunkts und der Anforderungen an die Meldepflicht erkennen.

In der deutschen Rechtsordnung wird die zuständige Marktüberwachungsbehörde erst nach der unternehmensinternen Risikobewertung des potenziellen Produktfehlers und der Entscheidung über Korrekturmaßnahmen unterrichtet. Es besteht die gesetzliche Pflicht, dass Unternehmen die zuständige Behörde informieren müssen, wenn sie wissen, dass ein Produkt ein Risiko für die Gesundheit und Sicherheit für Personen darstellt (vgl. ProdSG §6 Abs.4). Obwohl nicht ausdrücklich formuliert, gilt für die deutsche Meldepflicht allerdings der Verhältnismäßigkeitsgrundsatz. Danach kann es im Einzelfall zum Entfallen der Meldepflicht kommen, wenn lediglich Anhaltspunkte vorliegen, aber noch keine Gewissheit besteht, ob vom Produkt tatsächlich eine Gefahr ausgeht. Die Beschränkung auf Meldungen von kritischen Fällen ist laut des Gesetzgebers damit zu rechtfertigen, dass dadurch eine übermäßige Belastung der Behörden vermieden werden soll (vgl. Wilrich, 2012, S.178). Somit zeigt sich in der deutschen Rückrufpraxis, dass Behörden in der Regel nur dann informiert werden, wenn sich ein Unternehmen bereits für einen Rückruf entschieden hat (vgl. Foodwatch, 2017, S.21).

Hier liegt ein wesentlicher Unterschied zum amerikanischen Gesetz, welches von Unternehmen eine wesentlich frühere Unterrichtung der CPSC unter strengeren Bedingungen verlangt. Sobald ein Unternehmen bereits Anhaltspunkte darüber hat, dass von einem Produktfehler eine Gefahr (substantial product hazard) ausgehen könnte, muss es die CPSC umgehend informieren (vgl. CPSC, 2012, S.8). Die Entscheidung für einen Rückruf und weitere Korrekturmaßnahmen werden erst nach der Unterrichtung in Einvernehmen mit der CPSC entschieden (vgl. CPSC, 2012, S.17). Im Vergleich zu den recht moderaten Sanktionen in der deutschen Rechtsordnung, die drohen, wenn ein Unternehmen die Meldepflicht verletzt, scheut die CPSC nicht davor schwerwiegende Strafen für Unternehmen auszusprechen (vgl. CPSC, 2013a, S.11 , §39 Abs.1 Art.4 i.V.m. Abs.2 ProdSG).) Möglich wäre, dass die harten Strafen in den USA im Fall, dass Unternehmen ihre Meldepflicht nicht in angemessener Weise Folge leisten, zum einen die zentrale Rolle der CPSC im gesamten Rückrufprozess verdeutlichen, als auch eine abschreckende Wirkung erzielen, um Unternehmen hinsichtlich der Wichtigkeit eines angemessenen Umgangs mit Produktfehlern zu sensibilisieren.

5.2.2. Entscheidung über Korrekturmaßnahmen

Die Entscheidung über Korrekturmaßnahmen umfasst die Wahl, ob im Falle eines Produktfehlers ein Rückruf, eine Warnung oder eine Rücknahme eingeleitet wird. Darüber hinaus beinhalten die Korrekturmaßnahmen die angebotene Kompensation sowie das Zusammenstellen eines Kommunikationsprogramms (vgl. CPSC, 2012, S.17).

Wie bereits erläutert hängt die Entscheidung über Korrekturmaßnahmen von der Effizienz der vorhandenen Rückverfolgbarkeit ab. Falls ein Unternehmen gewährleisten kann,

dass ihnen alle Kontaktdaten der betroffenen Konsumenten zur Verfügung stehen, können sie den Aufwand für die Nutzung der öffentlichen Massenmedien reduzieren oder sogar vermeiden. In diesen stillen Rückruf können betroffene Produktbenutzer unter Ausschluss der Öffentlichkeit direkt über die Rückrufaktion informiert werden (vgl. Klindt und Wende, 2014, S.86). Ist eine direkte Kommunikation mit den Konsumenten nicht möglich ist ein öffentlicher Rückruf notwendig, um betroffene Konsumenten zu erreichen (vgl. CPSC, 2012, S.20).

In der Entscheidung über Korrekturmaßnahmen kommt es zu einer gänzlich unterschiedlichen Regulierungsdichte in den Rechtsordnungen. Insbesondere über die Frage des Zeitpunkts der Entscheidung und der Entscheidungsgewalt der Rückrufmaßnahmen bestehen Unterschiede zwischen den Ländern. Eine Übereinstimmung besteht insofern darin, dass beide Rechtsansichten der Ansicht sind, dass eine persönliche Kontaktaufnahme zu betroffenen Verbrauchern im Idealfall genutzt werden sollte. Falls eine individuelle Kommunikation mit betroffenen Produktbesitzern nicht möglich ist, wird ein öffentlicher Rückruf in beiden Rechtsordnungen als unvermeidbar angesehen (vgl. ProSafe, 2011, S.24, CPSC, 2003, S.8).

In Deutschland entscheidet ein Unternehmen zunächst selbst, welche Korrekturmaßnahmen es wählt. Die zuständige Marktüberwachungsbehörde wird erst dann benachrichtigt, wenn sich ein Unternehmen für einen Rückruf entschieden hat (vgl. Foodwatch, 2017, S.21). Im Allgemeinen erarbeitet das Unternehmen erst einen Plan über Korrekturmaßnahmen, welcher anschließend von der zuständigen Behörde hinsichtlich seiner Angemessenheit zur Gefahrenabwehr bewertet wird (vgl. Anhang B). Hingegen werden Korrekturmaßnahmen in den USA erst nach der behördlichen Meldung eines Produktfehlers im Einvernehmen mit der CPSC in einem corrective action plan erarbeitet. So nimmt die CSPC in den USA im gesamten Entscheidungsprozess eine zentrale Rolle ein (vgl. CPSC, 2012, S.17).

Bei der Entscheidung über Korrekturmaßnahmen hat ein Unternehmen in Deutschland beträchtliche Entscheidungsspielräume. Es gibt keine gesetzlichen Vorgaben, welche Kommunikationskanäle für eine Rückrufaktion genutzt werden müssen (vgl. Foodwatch, 2017, S.86). Theoretisch stehen den deutschen Behörden Befugnisse zu, in die Ausgestaltung des Rückrufs einzugreifen (vgl. Anhang B). In der Rückrufpraxis scheint diese Befugnis gegenüber den Unternehmen aber offenbar begrenzt (vgl. Foodwatch, 2017, S.57). Meistens legen Unternehmen ihren Plan über Korrekturmaßnahmen vor und die Behörde stimmt im Wesentlichen zu. Laut der Verbraucherschutzorganisation Foodwatch (2017) kommt es in der Praxis nur selten vor, dass eine Behörde mit den ergriffenen Maßnahmen nicht einverstanden ist. Wenn der Fall eintritt, dass eine Behörde Anordnungen erlässt, beschränken sich diese oftmals auf die Entscheidung, ob ein Rückruf erfolgen muss (vgl. Foodwatch, 2017, S.57). Eine Anordnung, die Unternehmen zur zusätzlichen Nutzung von Kommunikationskanälen oder zu speziellen Formulierungen verpflichtet, erfolgt in der deutschen Rückrufpraxis wohl nicht einheitlich, denn viele Unternehmen klagen über unterschiedliche Bewertungsmaßstäbe von Behörden aufgrund von fehlenden einheitlichen Verwaltungsvorschriften (vgl. Foodwatch, 2017, S.58). Der Grund liegt darin, dass die Marktüberwachung in Deutschland föderal organisiert ist (BMUB, 2016, S.1) Somit kann angenommen werden, dass es zwischen den unterschiedlichen Marktüberwachungsbehörden durchaus zu Divergenzen in der Verteilung von fachlichen und personellen Kompetenzen kommt und Rückrufe je nach Behörde unterschiedlich geregelt werden (vgl. Foodwatch, 2017, S.55). So bestätigt ein Vertreter des MULNV NRW, dass die Behörde die vom Unternehmen gewählten Informationskanäle sowie die Formulierungen in der Rückrufanzeige durchaus geprüft und Nachforderungen bei unzureichenden Korrekturmaßnahmen angeordnet (vgl. Anhang B). Hingegen hat Foodwatch (2017) in einer aktuellen Studie rund 90 Produktrückrufe in Deutschland ausgewertet und festgestellt, dass es zu großen Unterschieden hinsichtlich der Nutzung von Informationskanälen zur Bekanntgabe eines Rückrufs kommt (vgl. Foodwatch, 2017, S.86).

Hier liegt ein entscheidender Unterschied zur US-amerikanischen Rechtsordnung, in der die CPSC als unabhängige bundestaatliche Verbraucherschutzbehörde für alle meldepflichtigen Rückrufe von Verbraucherprodukten im gesamten US-Raum zuständig ist (vgl. Haberland, 2016, S.148). Ferner wird US-amerikanischen Unternehmen bei der Ausgestaltung eines Rückrufs ein wesentlich kleinerer Spielraum bei der Entscheidung für Korrekturmaßnahmen zugestanden, als es in Deutschland der Fall ist. Korrekturmaßnahmen werden vom betroffenen Unternehmen gemeinschaftlich mit der CPSC festgelegt und in einem konkreten Rückrufplan (corrective action plan) ausgearbeitet (vgl. CPSC, 2012, S.17). Während deutsche Behörden unterschiedliche Anforderungen an die Nutzung von Kommunikationskanälen für einen öffentlichen Rückruf stellen, gibt die CPSC je nach dem beurteilten Schweregrad des Produktfehlers einheitliche Vorgaben hinsichtlich der Art und des Umfangs der zu nutzenden Kanäle. Je höher die Gefahr eingeschätzt wird, die vom betroffenen Produkt ausgeht, desto umfangreicher sind die Vorgaben zur Rückrufkommunikation (vgl. CPSC, 2012, S.14).

Ein weiterer Unterscheidungsaspekt ist die Formulierung und Ausgestaltung der Rückrufanzeige. So kritisiert Klaus Müller, Vorstand der Verbraucherschutzzentrale Bundesverband (vzbz) und Deutschlands oberster Verbraucherschützer, dass deutsche Unternehmen in der Regel oft eine verharmlosende Wortwahl benutzen und Gesundheitsrisiken unverständlich formulieren (vgl. Anhang A: Experteninterview mit Klaus Müller).

Es gibt keine gesetzlichen Vorgaben, wie Unternehmen in Deutschland eine Rückrufanzeige formulieren müssen. Ob potenzielle Gefahren präzise benannt oder nur vage Angaben darübergemacht werden, ob ein Foto des Produktes veröffentlicht wird – bei all diesen Aspekten verfügen deutsche Unternehmen über große Entscheidungsspielräume (vgl. Foodwatch, 2017, S. 21). Dagegen bestehen in den USA gesetz-

liche Vorgaben über den Inhalt und der Formulierung von Rückrufmitteilungen. So muss in der Rückrufanzeige ein Foto des unsicheren Produktes abgebildet und Angaben über die Anzahl und Art der mit dem Produkt verbundenen Verletzungen und Todesfälle gemacht werden (vgl. 15.U.S.C. § 2064 (i)). Im Falle eines öffentlichen Rückrufs wird eine gemeinschaftlich formulierte Pressemitteilung nach den gesetzlichen Anforderungen von der CPSC und dem Unternehmen veröffentlicht. Bevor eine Rückrufmitteilung an die Öffentlichkeit gelangt, brauchen sowohl das Unternehmen als auch die CPSC die Genehmigung des jeweils anderen (vgl. CPSC, 2012, S.20). In der US-amerikanischen Rückrufpraxis zeigt sich, dass praktisch jedes Wort, das in Rückrufanzeigen verwendet wird, ausdiskutiert wird (vgl. Felcher, 2003, S. 178).

Wie bereits erläutert, besteht ein wichtiger Bestandteil der Korrekturmaßnahmen in der Entscheidung, welche Kompensation den betroffenen Konsumenten angeboten wird. Hiermit sind angebotene Ausgleichsleistungen von Unternehmen gemeint (vgl. Grunwald, 2008, S.47). In beiden Rechtsordnungen werden zwar vergleichbare Vorschläge über eine mögliche Kompensationen gemacht, allerdings ergibt sich auch hier der Unterschied, dass die Wahl der Kompensation in den USA gemeinschaftlich zwischen der CPSC und dem Unternehmen entschieden wird, während in Deutschland das Unternehmen den Hauptverantwortlichen in der Entscheidung darstellt (vgl. CPSC, 2012, S.17). Darüber hinaus besteht in der deutschen Rechtsordnung keine gesetzliche Kostentragungspflicht für Reparaturmaßnahmen, während die Rechtslage in den USA vom Unternehmen eine Übernahme aller anfallenden Kosten verlangt (vgl. Haberland, 2016, S.196). Zwar entscheiden sich deutsche Unternehmen aufgrund drohender Reputationsschäden in den meisten Fällen für einen kostenfreien Rückruf, dennoch liegt dies im Ermessen des Unternehmens selbst und ist in der deutschen Rechtsordnung nicht gesetzlich verpflichtend (vgl. Klindt und Wende, 2014, S.85).

5.2.3. Durchführung, Überwachung und Beendigung des Rückrufs

Evidente Unterschiede sind ebenfalls im letzten Schritt des Rückrufmanagements zwischen der deutschen und der US-amerikanischen Rechtsordnung zu verzeichnen. Insbesondere hinsichtlich der behördlichen Kontrolldichte über den Erfolgsgrad eines Produktrückrufs und der Entscheidungsbefugnis über die Beendigung des Rückrufs entstehen länderspezifische Differenzen. Die Durchführung der Maßnahmen umfasst zum einen die Sicherstellung, dass der Verkauf, Import oder die Distribution der betroffenen Produkte gestoppt wird und erfolgreich aus der Lieferkette entfernt werden. Zum anderen muss ein Unternehmen sicherstellen, dass die betroffenen Produkte, die sich überwiegend bei Verbrauchern befinden, durch die getroffenen Maßnahmen erreicht und zur Teilnahme am Rückruf motiviert werden (vgl. ProSafe, 2011, S.13). Ob ein Unternehmen die gewählten Korrekturmaßnahmen realisiert wird in den beiden Rechtsordnungen unterschiedlich überwacht. In Deutschland ist ein Unternehmen nicht verpflichtet Entwicklungen im Rückruf, wie die Rücklaufquotenentwicklung, zu dokumentieren oder der Marktüberwachungsbehörde zu melden. Somit wird von der zuständigen Behörde auch nicht kontrolliert, ob ein Unternehmen eine hohe Rücklaufquote erreicht (vgl. Anhang B).

Die CPSC hingegen überwacht die unternehmerische Umsetzung der Korrekturmaßnahmen aktiv durch regelmäßige Kontrollbesuche. Während deutsche Unternehmen keiner Dokumentationspflicht über die Entwicklung des Rückrufs unterworfen sind, müssen US-amerikanische Unternehmen eine monatliche Berichterstattung über die Rücklaufquotenentwicklung an die CPSC leisten (vgl. CPSC, 2012, S.26). Ob ein Rückruf offiziell beendet wird, hängt von einer abschließenden Bewertung der CPSC ab. Falls die US-Behörde nicht der Ansicht ist, dass der Rückruf ergebnisreich genug war, kann sie die Durchführung von umfassenderen Korrekturmaßnahmen anordnen (vgl. ebd.). Somit zeigt sich, dass die CPSC entscheidet, wann ein Rückruf beendet werden darf und wann weitere Korrekturmaßnahmen erforderlich sind. Dagegen ist in Deutschland das Unternehmen der alleinige Entscheidungsträger, wann ein Rückruf beendet wird, da deutsche Behörden in der Regel keine zweite Durchführung von Korrekturmaßnahmen anordnen (vgl. Anhang B).

Final zeigt sich, dass es in der Rückrufregelung zu großen Unterschieden zwischen der deutschen und der US-amerikanischen Rechtsordnung kommt. Die Gegenüberstellung verdeutlicht, dass die Regulierungsdichte in Bezug auf Rückrufe von Verbraucherprodukten in den USA deutlich höher und einheitlicher ist als in Deutschland. Darüber hinaus lässt sich erkennen, dass die Stellung der Behörde im gesamten Rückrufprozess in den USA viel stärker ist als im deutschen föderalen System.

6. Einflussfaktoren auf das Konsumentenverhalten in Rückrufsituationen

Ohne die aktive Teilnahme am Rückruf von betroffenen Konsumenten bleiben jegliche Bemühungen für ein erfolgreiches Rückrufmanagement erfolglos. Folglich muss die Rücklaufquotenentwicklung ebenfalls als verhaltenswissenschaftliche Herausforderung angesehen werden (vgl. Van Venrooy, 2015, S.19). Erklärtes Ziel dieses Kapitels ist es Erkenntnisse aus der Rückrufforschung zu dem interessierenden Themengebiet der Produktrückrufe zu erschließen, wobei allerdings bisher nur wenige Studien darüber existieren. Es sollen Faktoren aufgezeigt werden, die eine stärkere Einbindung der betroffenen Konsumenten in den Rückrufprozess erlauben und Unternehmen optimierte Beiträge zur Rücklaufquotenentwicklung ermöglichen. Einen Überblick über mögliche Einflussfaktoren auf die Teilnahme an einem Produktrückruf vermittelt Tabelle 2. Sie lassen sich nach, kommunikations-, verständnis- und motivationsrelevanten Einflussfaktoren systematisieren.

Kommunikationsrelevante Einflussfaktoren

Die jeweilige Kontaktebene mit betroffenen Konsumenten stellt einen wichtigen Aspekt für die Übermittlung der Rückrufaufforderung dar (vgl. Van Venrooy, 2015, S.346). Somit

Tabelle 2: Einflussfaktoren auf die Teilnahme am Produktrückruf (Quelle: eigene Darstellung)

Einflussfaktoren	Studie	Einfluss	Kommentar
kommunikationsrelevant			
direkte Kommunikation	Murphy und Rubin (1988)	+	signifikant positiver Einfluss auf den Rückruferfolg
	NHTSA (2002)		
verständnisrelevant			
unverständliche Formulierung der Rückrufinformation	Van Venrooy (2015)	-	eine verständliche Formulierung wird als Kernvariable für die Rückrufteilnahme charakterisiert
technokratische Formulierung zur Beschreibung des Produktfehlers	Van Venrooy (2015)	-	Negativer Einfluss auf die Entscheidung zur Rückrufteilnahme
Motivationsrelevant			
Teilnahmekosten	Murphy und Rubin (1988)	-	Hohe Teilnahmekosten haben einen negativen Einfluss auf die Teilnahmeentscheidung
Nutzen der Rückrufteilnahme			
- vollstände Kompensation (full remedy)	ableitend aus: Kelley et al. (1993)	+	Eine vollständige Kompensation stellt für Konsumenten ein attraktiveres Angebot dar als eine partielle Kompensation (partial remedy)
- zusätzliche monetäre Anreize	Van Venrooy (2015)	+	positiver Einfluss auf die Entscheidung zur Rückrufteilnahme

erweist sich die Art der Rückrufkommunikation als ein Einflussfaktor auf die Entscheidung zur Rückrufteilnahme von Konsumenten. Eine Regressionsanalyse von Murphy und Rubin (1988) zeigt, dass eine direkte, individuelle Kommunikation mit betroffenen Konsumenten einen signifikant positiven Einfluss auf den Rückruferfolg hat (vgl. Murphy und Rubin, 1988, S. 21). Diese Erkenntnis wird ebenfalls durch eine aktuelle Studie der NHTSA bestätigt, in der aufgezeigt wird, dass seit der gesetzlichen Verpflichtung zur Benutzung von Produktregistrierungskarten bei Produkten für Kleinkinder in den USA, welche eine direkte Rückrufkommunikation mit betroffenen Konsumenten ermöglichen, die durchschnittliche Rücklaufquote um sieben Prozent gestiegen ist (vgl. NHTSA, 2002 , S.16).

Verständnisrelevante Einflussfaktoren

Nicht nur die Rückrufkommunikation als solches, sondern auch das Verständnis der Rückrufkommunikation für die von einem Rückruf betroffenen Rezipienten ist ein zwingend notwendiges Element, das im Zuge eines erfolgreichen Produktrückrufs berücksichtigt werden muss (vgl. Van Venrooy, 2015, S.184). Somit stellt die Verständlichkeit der Information in einer Rückrufanzeige eine Kernvariable für die Entscheidungsfindung zur Teilnahme an einem Produktrückruf dar. Van Venrooy (2015) zeigt auf, dass die Formulierung der Rückrufanzeigen von Konsumenten häufig als zu unkonkret und zu unverständlich empfunden wird. Werden Rückrufanzeigen nicht eindeutig und einfach verständlich formuliert, kommt es unweigerlich zu Problematiken, die sich in der Phase der Entscheidungsfindung bei Konsumenten negativ auswirken. Eine verweigerte Teilnahme der Betroffenen ist die Konsequenz dieser Situation (vgl. Van Venrooy, 2015, S.318). Auch die Verwendung von technischen Fachbegriffen zur Erklärung des Produktfehlers wird für die Rückrufkommunikation an technischen Laien als ungeeignet und kompliziert angesehen. Die in einer Rückrufanzeige zur Verfügung gestellten Informationen beschreiben Konsumenten oft als zu technokratisch, um die Problematik des betroffenen Produkts zu verstehen. Dies wirkt sich folglich negativ auf die Entscheidung zur Rückrufteilnahme aus (vgl. Van Venrooy, 2015, S. 275).

Motivationsrelevante Einflussfaktoren

Eine zwingende Voraussetzung für eine maximale Rücklaufquotengenerierung ist die unternehmerische Fähigkeit, die betroffenen Konsumenten zur aktiven Teilnahme am Rückruf zu motivieren (vgl. Van Venrooy, 2015, S.350).

Für die Motivation zur Teilnahme spielt das von Konsumenten subjektiv wahrgenommene Kosten-Nutzen Verhältnis einer Rückrufteilnahme eine Rolle (Harvard Business Review, 1996, S.16). So stellt die Art der angebotenen Kompensation einen motivational wirkenden Faktor dar, der die Entscheidung von Konsumenten zur Teilnahme beeinflusst. Allgemein zeigt sich, dass Konsumenten eine vollständige Kompensation (kostenloser Austausch, vollständige Kaufpreisrückerstattung oder kostenlose Reparatur) als eine attraktivere Option ansehen als eine partielle Kompensation (Ermäßigungen für zukünftige Käufe oder Reperatursets zur eigenständigen Reparatur) (vgl. Kelley et al., 1993, S. 438 ff.). Liu et al. (2016) schließen daraus, dass die Rückrufteilnahme höher ausfal-

len wird, wenn Unternehmen eine vollständige Kompensation anbieten. (vgl. Liu et al., 2016, S.82). Auch das Angebot von zusätzlichen monetären Anreizen kann einen positiven Einfluss darauf haben, ob Konsumenten an einem Rückruf teilnehmen. Gemäß Van Venrooy (2015) können bereits kleine Anreize zur Rückrufteilnahme, wie die Ausstellung von Gutscheinen, eine deutlich positive Wirkung auf die Rückrufteilnahme entfalten (vgl. Van Venrooy, 2015, S.353).

Da ein Produktrückruf für Konsumenten in der Regel Kosten verursachen kann, geht die subjektive Wahrnehmung der Teilnahmekosten in die Entscheidungsfindung zur Teilnahme am Rückruf ein (vgl. Van Venrooy, 2015, S.319f.). Zu denken sind bei Teilnahmekosten an zeitliche, physische und finanzielle Aufwände, wie Verpackungsmaterial für Produkte zu finden und diese zur Post oder zum Händler zu bringen oder finanzielle Aufwände für Porto oder Fahrtkosten (vgl. Bodewig, 1999, S.95). Kosten für die Rückrufteilnahme stellen einen Faktor dar, der sich negativ auf die Entscheidung zu einer Rückrufteilnahme auswirkt (vgl. CPSC, 2003, S.23). So zeigt eine Untersuchung der CPSC, dass bereits geringe Unannehmlichkeiten einen signifikanten Einfluss auf die Rückrufteilnahme ausüben (CPSC, 2003, S.23). Dass die Umstände, die mit einer Produktrückgabe zum Hersteller oder Händler verbunden sind, häufig als beschwerlich empfunden werden und einen negativen Einfluss auf die Entscheidung zur Rückrufteilnahme ausüben, wird auch von Murphy und Rubin (1988) betont (vgl. Murphy und Rubin, 1988, S. 24 f.)

Aus diesen Forschungsergebnissen lässt sich ableiten, dass je geringer der für die Rückrufteilnahme erforderliche physische, zeitliche und finanzielle Aufwand von Konsumenten wahrgenommen wird und je höher der mit einer Rückrufteilnahme verbundene Nutzen (in Form einer vollständigen Kompensation und zusätzlichen Anreizen) ausfällt, desto eher sind Konsumenten motiviert am Rückruf teilzunehmen.

7. Handlungsempfehlungen für die Rückrufpraxis

In dem abschließenden Kapitel sollen aus den aufgezeigten Unterschieden in der Regelung eines Rückrufmanagements zwischen den Rechtsordnungen und den herausgearbeiteten Einflussfaktoren auf die Teilnahme am Rückruf Handlungsempfehlungen für die Rückrufpraxis abgeleitet werden, die eine Optimierung der Rücklaufquoten ermöglichen.

Die Sicherstellung einer verständlichen und motivierenden Aufforderung zum Rückruf

Unternehmen müssen in einer Rückrufsituation sicherstellen, dass alle betroffenen Konsumenten durch gewählte Maßnahmen erreicht und zu einer Teilnahme am Rückruf motiviert werden. Nur unter dieser Voraussetzung kann eine maximale Rücklaufquote erreicht werden (vgl. Van Venrooy, 2015, S.71). Falls die direkte Kommunikation mit betroffenen Konsumenten nicht möglich ist, muss ein Unternehmen angemessene Kommunikationswege wählen um sicherzustellen, dass alle betroffenen Konsumenten über den Produktrückruf informiert werden. Um Konsumenten überhaupt effektiv und gezielt ansprechen zu können, müssen Unternehmen auch die Voraussetzung für eine Verständlichkeit der Rückrufaufforderung schaffen. Wie die empirischen Untersuchungen aufzeigen, wirkt sich eine unverständliche und zu technokratische Formulierung negativ auf die Entscheidung der Konsumenten zur Teilnahme am Rückruf aus (vgl. Van Venrooy, 2015, S. 275). Diese Verständnisproblematik wird von vzbz-Vorstand Klaus Müller in der deutschen Rückrufpraxis beobachtet: „Wir erleben in Deutschland eher verharmlosende Wortwahlen über Risiken und unzureichende Formulierungen. Dabei ist der Verbraucherschutz der Meinung, dass man sich am Durchschnittsverbraucher orientieren sollte (...). Dagegen geht die Industrie eher von einem mündigen, gut informierten Verbrauchern aus und unterstellt damit regelmäßig von Verbrauchern fachliches Wissen, die man ihnen jedoch nicht unterstellen kann.“(Anhang A) Damit keine Verständnisproblematiken aufkommen, müssen Formulierungen so gewählt werden, dass Informationen über den Rückruf und dem Produktfehler eindeutig und auch für technische Laien verständlich sind (vgl. Van Venrooy, 2015, S.349).

Um Konsumenten zu einer Teilnahme am Rückruf zu motivieren und die Rücklaufquotenentwicklung positiv zu beeinflussen, zeigen Studien aus der Rückrufforschung, dass Unternehmen Konsumenten eine vollständige Kompensation und monetäre Zusatzanreize anbieten sollten (vgl. Kelley et al., 1993, S.438ff., Van Venrooy, 2015, S. 353). Gleichzeitig muss ein Unternehmen sicherstellen, dass die Teilnahme am Rückruf für Konsumenten mit keinen Kosten verbunden ist. Somit ergibt sich für deutsche Unternehmen die direkte Handlungsempfehlung, trotz der fehlenden Kostentragungspflicht für Reparaturleistungen, die mit einer Teilnahme am Rückruf verbundenen Kosten für Konsumenten zu übernehmen Zwar bedeutet dies für die Unternehmen einen Anstieg der direkten Kosten, allerdings kann somit die Entwicklung der Rücklaufquote positiv beeinflusst werden (vgl. Liu et al., 2016, S.92).

Rechtliche Vorgaben präzisieren und einheitliche Kontrollen durch Behörden sicherstellen

Natürlich wollen Unternehmen im Falle des eigenen Produktversagens sicherstellen, dass alle unsicheren Produkte erfolgreich entfernt werden, denn der Schutz der Verbraucher hat stets oberste Priorität für Unternehmen (vgl. Grocery Manufacturers Association, 2014, S.2). Betrachtet man die Rückrufpraxis, so kann allerdings nicht von der Hand gewiesen werden, dass sich Unternehmen in der Entscheidung über Korrekturmaßnahmen manchmal in einem Interessenkonflikt zwischen Verbraucherschutz und wirtschaftlichen Überlegungen befinden, da sie Imageverluste und hohe Kosten im einem Rückruf möglichst vermeiden möchten (vgl. Foodwatch, 2017, S.20). So zeigt sich in den USA, dass in der gemeinsamen Ausgestaltung des Rückrufplans nahezu jedes Wort in Rückrufanzeigen zwischen der CPSC und dem Unternehmen ausdiskutiert wird (vgl. Felcher, 2003, S.178) In Deutschland klagt der Verbraucherschutz über unzureichende Formulierungen in der Rückrufkommunikation (vgl. Anhang A). Ob dies ungewollt aus Fehleinschätzungen oder

bewusst aus Angst vor Imageverlusten geschieht, kann an dieser Stelle nicht beantwortet werden. Dennoch zeigt der Vergleich der Rückrufregelungen, dass es in der deutschen Rechtsordnung im Vergleich zur US-amerikanischen Rückrufpraxis an gesetzlichen Vorgaben zur Ausgestaltung von Rückrufen und an einer einheitlichen behördlichen Überwachung der Durchführung von Rückrufen fehlt. So fordert Klaus Müller: „Bei Rückrufen von Verbraucherprodukten brauchen wir in Deutschland erstens Standards, die das eindeutig definieren und zweitens ebenfalls auch Überprüfungen durch die zuständigen Behörden“ (vgl. Anhang A).

Zwar sind einheitliche Standards zur Ausgestaltung eines Rückrufs und behördliche Überwachungen wichtig, allerdings müssen Korrekturmaßnahmen im Einzelfall auch hinsichtlich ihrer Angemessenheit, Erforderlichkeit und Zumutbarkeit abgewogen werden (vgl. Bodewig, 1999, S.13). Denn in einigen Rückrufsituationen mögen Abwägungen über Kommunikationskanäle und Formulierungen von Unternehmen zulässige Überlegungen sein. Ruft ein Unternehmen ein Produkt zurück, das ein niedriges Risiko unter einer sehr geringen Wahrscheinlichkeit darstellen könnte, ist es fraglich ob es eigene Produktfehler in einem hohen Grad medial auf den eigenen Social-Media-Kanälen verbreiten muss und dadurch potenzielle Imageschäden und erhebliche Umsatzeinbußen zu befürchten hat. Stellt aber ein zurückgerufenes Produkt eine wirkliche Gefahr für Konsumenten dar, muss von Behörden einheitlich sichergestellt werden, dass Unternehmen alles in ihrer Macht Stehende tun, um alle unsicheren Produkte aus dem Markt zu entfernen. In einer solchen Situation erscheint es umso wichtiger, durch Vorgaben zur Ausgestaltung des Rückrufs zu garantieren, dass verständlich und in einem angemessenen Umfang kommuniziert wird, um ein zweifelfreies Empfangen und Verständnis der betroffenen Konsumenten sicherstellen zu können (vgl. Van Venrooy, 2015, S. 184) Hier wäre es in Deutschland wichtig einheitliche Bewertungsmaßstäbe für die Marktüberwachungsbehörden in den Bundesländern durch eine allgemeine Verwaltungsvorschrift zu schaffen oder sich am Modell der USA zu orientieren, in dem die CPSC als einheitliche bundesstaatliche Verbraucherschutzbehörde agiert. Gerade hinsichtlich der Tatsache, dass ein zunehmend größerer Teil der Produktionen im Rahmen globaler Wertschöpfungsketten stattfinden „sind regionale Marktüberwachungsbehörden wie in Deutschland im 21. Jahrhundert nicht mehr zeitgemäß“ (vgl. Anhang A).

Die Rückverfolgbarkeit als Rückgrat eines erfolgreichen Produktrückrufs

Wie sich in der Gegenüberstellung der Regelung eines Rückrufmanagements in der US-amerikanischen und deutschen Rechtsordnung gezeigt hat, ist jedes Unternehmen sowohl in Deutschland als auch in den USA selber dafür verantwortlich präventive Maßnahmen zur Sicherstellung der Rückverfolgbarkeit von Produkten und deren Besitzern zu ergreifen. Betrachtet man den gesamten Prozess eines Rückrufmanagements in Deutschland und in den USA, so stellt die Sicherstellung der Rückverfolgbarkeit das eigentliche Rückgrat eines zielgerichteten und erfolgreichen Produktrückrufs und folglich auch einer Maximierung der Rücklaufquote dar. Denn je genauer zugeordnet werden kann, in welchen Produkten das jeweilige Sicherheitsproblem liegt, wo sich diese Produkte befinden und welche Konsumenten dadurch betroffen sind, desto eingegrenzter und präziser kann ein Rückruf durchgeführt werden (vgl. Klindt und Wende, 2014, S87). Folglich sollten Unternehmen aus eigener Initiative gezielt präventive Maßnahmen zur Sicherstellung der Rückverfolgbarkeit ergreifen. Eine Möglichkeit zur Optimierung der Produktrückverfolgbarkeit für Konsum- und Gebrauchsgüterhersteller stellt hier das Efficient Consumer Response System (ECR) dar. Dabei handelt es sich um eine Kooperation von Handel- und Herstellerunternehmen, welche strategische Partnerschaften bilden, um gemeinsam Prozessschritte zu entwickeln, die eine Produktrückverfolgbarkeit und Rückrufaktionen erleichtern sollen (vgl. Centrale for Coorganisation GmbH, 2004, S.6 ff.).

Die Individualisierung der Rückrufkommunikation als Best Practice eines Produktrückrufs

Nicht nur die Effizienz der Sicherstellung der Produktrückverfolgbarkeit in der Lieferkette, sondern auch die Rückverfolgbarkeit von Kunden ist für die Erreichung einer maximalen Rücklaufquote von hoher Bedeutung, da sie eine direkte Kommunikation mit betroffenen Konsumenten sicherstellt, welche einen signifikanten positiven Einfluss auf die Entscheidung der Konsumenten zur Rückrufteilnahme hat (vgl. Murphy und Rubin, 1988, S. 23).

Betrachtet man die Interessenslagen der Akteure einer Rückrufaktion, also das Unternehmen, die Behörde und die betroffenen Konsumenten, so kann die direkte Rückrufkommunikation durchaus als Best Practice eines Produktrückrufs angesehen werden. Aus Sicht der Behörde kann eine direkte Rückrufkommunikation mit betroffenen Konsumenten positiv bewertet werden, da diese eine Rückholung gefährlicher oder nicht konformer Produkte aus der Verbraucherhand positiv beeinflusst (vgl. CPSC, 2012, S.5). Die individuelle Rückrufkommunikation ist auch aus der Unternehmensperspektive positiv zu bewerten. In erster Linie hinsichtlich der Optimierung des Verbraucherschutzes, aber auch dahingehend, dass die Gefahr vor Imageverlusten durch eine mediale Aufmerksamkeit über die eigenen Produktschwächen infolge eines öffentlichen Rückrufs vermieden werden kann. Kunden, die Produkte gekauft haben, von denen kein Fehler ausgeht, müssen so nicht unnötig beunruhigt werden. Für Konsumenten liegt eine individuelle Rückrufkommunikation insofern in ihrem Interesse, da sie so gezielt Rückrufinformationen bekommen, die die eigenen gekauften Produkte betreffen. Einen Vorteil der direkten Rückrufkommunikation gegenüber einem öffentlichen Rückruf sieht auch vzbz-Vorstand Klaus Müller: „Ich glaube die wenigsten Menschen lesen Produktrückrufe wirklich in Tageszeitungen oder sehen diese im Fernsehen. Und wenn wir ehrlich sind, wer guckt auf die behördlichen Internetportale und überprüft in einem gewissen Rhythmus Produktrückrufe (...). Das ist de facto lebensfremd. Und insofern glaube ich, dass erst dann, wenn

es eine Individualisierung gibt und in meinem E-Mail Korb ein Warnhinweis auftaucht, die Aufmerksamkeit bei Konsumenten erheblich höher ist.“ (vgl. Anhang A).

Allerdings setzt eine direkte Kontaktierung auch voraus, dass Unternehmen Kundendaten speichern und im Falle eines Rückrufs verwenden. Hierbei müsste nach Klaus Müller sichergestellt sein, dass der Datenschutz gewährleistet ist. Daten von Verbrauchern dürfen von Unternehmen nur dann in einem Produktrückruf verwendet werden, wenn Verbraucher einer Nutzung aktiv zugestimmt haben (vgl. Anhang A).

Aufgrund dieser Erkenntnis sollten Unternehmen in beiden Rechtsordnungen trotz mangelnder gesetzlicher Vorgaben aus eigener Initiative gezielt präventive Maßnahmen ergreifen um eine direkte Rückrufkommunikation mit betroffenen Konsumenten gewährleisten zu können.

Eine Überlegung wäre, Konsumenten Anreize für eine Registrierung zu bieten und ihnen aufzuzeigen, dass eine Registrierung durchaus in ihrem eigenen Interesse liegt, da sie bei potenziellen Produktwarnungen persönlich informiert werden könnten (vgl. CPSC, 2012, S.28).

Außerdem könnte die Kennzeichnung von Produkten beim Verkauf mit Kundendaten auf Verkaufsunterlagen, wie Rechnungen oder Garantiekarten, erfasst werden. Es ist anzunehmen, dass diese Methode in einem Fachhandel eher anwendbar ist, als beim Verkauf von Produkten des täglichen Bedarfs in einem Discounthandel (vgl. Klindt und Wende, 2014, S.87).

Darüber hinaus spielt der Online Handel bei der Ermöglichung einer Individualisierung der Rückrufkommunikation eine entscheidende Rolle. Der Online Händler weiß genau, welchem Konsumenten er ein vom Rückruf betroffenes Produkt ausgeliefert hat und könnte die betroffenen Konsumenten direkt per persönlicher E-Mail zum Rückruf auffordern. Auch hier ist aus datenschutzrechtlichen Gründen eine vorherige Einwilligung der Konsumenten notwendig (vgl. Anhang A). So verpflichtet sich Amazon bereits zur Unterstützung von Herstellern im Falle von Produktrückrufen (vgl. Amazon.com Inc, 2017). Da dies aber bisher noch kein Standard in der Rückrufpraxis ist, sollten Unternehmen, die ihre Produkte im höheren Umfang über den Online Handel veräußern, mit Online-Händlern eine Zusammenarbeit im Falle eines Rückrufs vertraglich festlegen (vgl. Anhang A).

8. Ausblick und Fazit

Die stetig steigende Anzahl an Produktrückrufe in Deutschland und auch weltweit wird die Bedeutung der vorliegenden Thematik weiter erhöhen. Komplexer werdende, globale Wertschöpfungsketten, die eine Produktrückverfolgbarkeit immer weiter erschweren und eine Erhöhung der gesellschaftlichen und rechtlichen Relevanz des Verbraucherschutzes zeigen, dass die Gestaltung von Rückrufen auch in Zukunft fordernd für Unternehmen wird. Man kann davon ausgehen, dass jedes verantwortungsvolle Unternehmen in seinen Unternehmensgrundsätzen fest verankert hat, seinen Kunden Produkte mit einem größtmöglichen Sicherheitsniveau anzubieten. Es kann allerdings nie komplett ausschließen, dass Fehler passieren. Sei es in der Entwicklung, Produktion oder in der Anwendung durch Verbraucher. Fehler passieren täglich und sie werden auch in Zukunft passieren. Die unzureichenden Rücklaufquoten bei Rückrufen von Verbraucherprodukten verdeutlichen allerdings die Notwendigkeit sich mit Möglichkeiten einer Optimierung der Durchsetzung von Produktrückrufen zu beschäftigen.

Die vorliegende Arbeit hat Maßnahmen vorgestellt, die ein Unternehmen präventiv und während eines Rückrufs ergreifen muss um alle betroffenen Produkte erfolgreich aus dem Markt zu entfernen. Es wurde aufgezeigt, dass Unternehmen eine Vielfalt an organisatorischen, rechtlichen und verhaltenswissenschaftlichen Einflussfaktoren berücksichtigen müssen, um eine vollständige Rückholung aller Produkte aus dem Markt zu erreichen. Aufgezeigte Unterschiede in der Regulierung von Rückrufen zwischen Deutschland und den USA erscheinen besonders vor dem Hintergrund der hohen deutschen Exportquote in die USA für eine erfolgreiche Durchführung von Rückrufen deutscher Exporteure von Bedeutung zu sein. Im Laufe der Untersuchungen stellte sich ebenfalls heraus, dass aus gesetzgeberischer Sicht durchaus noch Verbesserungspotenzial besteht, um sicherzustellen, dass Produktrückrufe in der deutschen Industrie verständlicher und letztlich damit sicherer durchgeführt werden.

Während sich die Rückrufforschung in den letzten Jahren besonders auf Auswirkungen von Rückrufen auf das Unternehmen, dessen Reputation und die Kundenwahrnehmung konzentriert hat, erscheint der Forschungsstand zu Einflussfaktoren auf das Konsumentenverhalten in einem Rückruf noch nahezu unerschlossen. Es scheint bisher noch wenig erforscht, warum sich Konsumenten zur Rückrufteilnahme entscheiden und warum es zu Kommunikations-, Verständnis- und Motivationsproblematiken in Produktrückrufen kommt. Hier bedarf es an weiteren Untersuchungen um die derzeitige Forschungslücke zu schließen.

Literatur

Amazon.com Inc. Rückrufe und Produktsicherheit, (Zugriff : 26.10.2017), 2017. URL https://www.amazon.de/gp/help/customer/display.html/?ie=UTF8&nodeId=200181490.

BAuA. „Produktrückrufe“, (Zugriff: 28.09.2017), 2017a. URL https://www.baua.de/DE/Themen/Anwendungssichere-Chemikalien-und-Produkte/Produktsicherheit/Rueckrufmanagement/Rueckrufmanagement.html.

BAuA. RAPEX - Rapid Exchange of Information System, (Zugriff: 16.09.2017), 2017b. URL https://www.baua.de/DE/Themen/Anwendungssichere-Chemikalien-und-Produkte/Produktsicherheit/Marktueberwachung/Rapex.html.

BAuA. Datenbank „Gefährliche Produkte in Deutschland“, (Zugriff: 30.10.2017), 2017c. URL https://www.baua.de/DE/Themen/Anwendungssichere-Chemikalien-und-Produkte/Produktsicherheit/Produktinformation/Datenbank/Produktsicherheit_form.html?meldev.GROUP=1&prodkat.GROUP=1.

BMUB. Marktüberwachung, (zugriff : 04..09.2017), 2016. URL http://www.bmub.bund.de/themen/wasser-abfall-boden/abfallwirtschaft/marktueberwachung/.

Bodewig, T. *Der Rückruf fehlerhafter Produkte: eine Untersuchung der Rückrufpflichten und Rückrufansprüche nach dem Recht Deutschlands, der Europäischen Union und der USA, Volume 36*. Mohr Siebeck, Tübingen, 1999.

BT-Drucksache 15/1629. Gesetzentwurf der Bundesregierung - Entwurf eines Gesetzes zur Einführung internationaler Rechnungslegungsstandards und zur Sicherung der Qualität der Abschlussprüfung (Bilanzrechtsreformgesetz - BilReG), 2003.

Bundesministerium für Arbeit, Soziales und Konsumentenschutz. Rückrufe von Verbraucherprodukten, (Zugriff : 17.08.2017), 2017. URL https://www.sozialministerium.at/cms/site/attachments/8/2/8/CH3434/CMS1490360582452/faq_rueckrufe.pdf.

Centrale for Coorganisation GmbH. Rückverfolgbarkeit von Produkten und effizienter Warenrückruf, (Zugriff:30.10.2017), 2004. URL https://www.gs1.at/fileadmin/user_upload/ECR_Rueckverfolgbarkeit_und_Warenrueckruf.pdf.

Coombs, W. T. und Holladay, S. J. *The Handbook of Crisis Communication*, volume 22. John Wiley & Sons, Chichester, 2011.

CPSC. Recall Effectiveness Research A Review and Summary of the Literature on Consumer Motivation and Behavior: USA, Berthesda (Zugriff : 30.09.2017), 2003. URL https://www.cpsc.gov/s3fs-public/pdfs/recalleffectiveness.pdf.

CPSC. Recall Handbook, (Zugriff : 28.09.2017), 2012. URL https://www.cpsc.gov/s3fs-public/8002.pdf.

CPSC. The regulated products Handbook, (Zugriff : 20.09.2017), 2013a. URL https://www.cpsc.gov/s3fs-public/RegulatedProductsHandbook.pdf.

CPSC. Consumer Registration Cards for Durable Infant or Toddler Products, (Zugriff: 30.09.2017), 2013b. URL https://www.cpsc.gov/Business--Manufacturing/Business-Education/Durable-Infant-or-Toddler-Products/Durable-Infant-or-Toddler-Product-Consumer-Registration-Cards.

CPSC. Staff Report to Congress Pursuant to Section 104(d) of the Consumer Product Safety Improvement Act of 2008, USA, Berthesda (Zugriff : 24.08.2017), 2016. URL https://cpsc.gov/s3fspublic/February_2016_Report_to_Congress_Pursuant_to_Section_104d_Recall_Notification_Effectiveness-FINAL.pdf.

Eilert, M., Jayachandran, S., Kalaignanam, K., und Swartz, T. A. Does it pay to recall your product early? An empirical investigation in the automobile industry. *Journal of Marketing*, 81(3):111–129, 2017.

Felcher, E. M. Product recalls: Gaping holes in the nation's product safety net. *The Journal of Consumer Affairs*, 37(1):170–179, 2003.

Foodwatch. Um Rückruf wird gebeten, (Zugriff: 27.10.2017), 2017. URL https://www.foodwatch.org/fileadmin/Themen/Rueckrufe/foodwatch_Report_Um_Rueckruf_wird_gebeten_final.pdf.

Gibson, D. C. Public relations considerations of consumer product recall. *Public Relations Review*, 21(3):225–240, 1995.

Glasl, T. und Klindt, T. *Krisenfall Produktrückrufe: erfolgreiches Management-Recht-Kommunikation*. Stuttgart, Boorberg, 2012.

Grocery Manufacturers Association. Succesfully Managing Produkt Recalls and Withdrawals, 4. Auflage, Washington DC, 2014.

Gröppel-Klein, A. und Germelmann, C. C. *Medien im Marketing: Optionen der Unternehmenskommunikation*. Wiesbaden, Springer-Verlag, 2009.

Grunwald, G. *Die Bewältigung von Produktkrisen: Ein Vergleich von Beschwerde-, Reklamations-und Rückrufsituationen*. Köln, Josef Eul Verlag, 2008.

Haberland, P. *Die Produkthaftung im deutschen und US-amerikanischen Recht*. Frankfurt, Peter Lang GmbH, 2016.

KBA. Kodex zur Ausführung des Produktsicherheitsgesetzes bei Straßenfahrzeugen, (Zugriff: 4.10.2017), 2011. URL https://www.kba.de/DE/Marktueberwachung/Kodex/kodex_pdf.pdf?__blob=publicationFile&v=4.

Kelley, S. W., Hoffman, K. D., und Davis, M. A. A typology of retail failures and recoveries. *Journal of Retailing*, 69(4):429–452, 1993.

Klindt, T. und Wende, S. *Rückrufmanagement: ein Leitfaden für die professionelle Abwicklung von Krisenfällen*. Berlin, 3. Auflage, Beuth Verlag, 2014.

Lach, S. und Polly, S. *Produktsicherheitsgesetz, Leitfaden für Hersteller und Händler*. Wiesbaden, 2. Auflage, Springer, 2015.

Lenz, T. *Produkthaftung*. München, Verlag C.H. Beck, 2014.

Liu, A. X., Liu, Y., und Luo, T. What drives a firm's choice of product recall remedy? The impact of remedy cost, product hazard, and the CEO. *Journal of Marketing*, 80(3):79–95, 2016.

Murphy, R. D. und Rubin, P. H. Determinants of recall success rates. *Journal of Products Liability*, 11(1):17–28, 1988.

NHTSA. Evaluation of Child Safety Seat Registration, DOT HS-809-518, Washington D.C., 2002.

NHTSA. U.S. Department of Transportation Launches New Public Awareness Campaign, (Zugriff: 1.10.2017), 2016. URL https://www.nhtsa.gov/press-releases/us-department-transportation-launches-new-public-awareness-campaign.

Oberlandesgericht Düsseldorf. Aktenzeichen I-15 U 129/14 , (Zugriff : 18.08.2017), 2015. URL https://www.ustiz.nrw.de/nrwe/olgs/duesseldorf/j2015/I_15_U_129_14_Urteil_20150910.html.

Pepels, W. *Produktmanagement*. 7. Auflage, Berlin, Duncker & Humblot GmbH, 2016.

ProSafe. Consumer Product Safety in Europe – Corrective Action Guide, (Zugriff:26.09.2017), 2011. URL http://www.prosafe.org/images/Documents/EMARS/Corrective_Action_Guide_Final-published.pdf.

Raithel, S. und Hock, S. J. *The Effects of Conforming and Nonconforming Product Recall Strategies on Customer and Investor Response*. Berlin, Fairfax, 2017.

Rettenbeck, S. *Die Rückrufpflicht in der Produkthaftung*. Baden-Baden, Nomos-Verlagsgesellschaft, 1994.

Rödl & Partner. *Handbuch internationale Produkthaftung*. 2. Auflage, Köln, Bundesanzeiger Verlag GmbH, 2014.

Tamme, A. *Rückrufkosten: Haftung und Versicherung*. Karlsruhe, Verlag Versicherungswirtschaft e.V., 1996.

Van Venrooy, M. *Produktrückrufe in der deutschen Automobilindustrie: Motivations-, Kommunikations-und Verständnisproblematiken*. Wiesbaden, Springer-Verlag, 2015.

Wilrich, T. *Das neue Produktsicherheitsgesetz (ProdSG)*. Berlin, Beuth Verlag, 2012.