JUMS

Volume 8, Issue 1, March 2023

JUNIOR MANAGEMENT SCIENCE

Advisory Editorial Board:

DOMINIK VAN AAKEN
FREDERIK AHLEMANN
JAN-PHILIPP AHRENS
BASTIAN AMBERG
THOMAS BAHLINGER
MARKUS BECKMANN
CHRISTOPH BODE
SULEIKA BORT
ROLF BRÜHL
KATRIN BURMEISTER-LAMP
JOACHIM BÜSCHKEN
CATHERINE CLEOPHAS
NILS CRASSELT
RALF ELSAS
KERSTIN FEHRE
MATTHIAS FINK
DAVID FLORYSIAK
GUNTHER FRIEDL
MARTIN FRIESL
FRANZ FUERST
WOLFGANG GÜTTEL
NINA KATRIN HANSEN
CHRISTIAN HOFMANN
SVEN HÖRNER
KATJA HUTTER
LUTZ JOHANNING
STEPHAN KAISER
NADINE KAMMERLANDER
ALFRED KIESER
NATALIA KLIEWER
DODO ZU KNYPHAUSEN-AUFSESS
SABINE T. KÖSZEGI
ARJAN KOZICA
CHRISTIAN KOZIOL
MARTIN KREEB
TOBIAS KRETSCHMER
WERNER KUNZ
HANS-ULRICH KÜPPER
MICHAEL MEYER
JÜRGEN MÜHLBACHER
GORDON MÜLLER-SEITZ
J. PETER MURMANN
ANDREAS OSTERMAIER
BURKHARD PEDELL
MARCEL PROKOPCZUK
TANJA RABL
SASCHA RAITHEL
NICOLE RATZINGER-SAKEL
ASTRID REICHEL
KATJA ROST
THOMAS RUSSACK
FLORIAN SAHLING
MARKO SARSTEDT
ANDREAS G. SCHERER
STEFAN SCHMID
UTE SCHMIEL
CHRISTIAN SCHMITZ
MARTIN SCHNEIDER
MARKUS SCHOLZ
LARS SCHWEIZER
DAVID SEIDL
THORSTEN SELLHORN
STEFAN SEURING
ANDREAS SUCHANEK
TILL TALAULICAR
ANN TANK
ORESTIS TERZIDIS
ANJA TUSCHKE
MATTHIAS UHL
CHRISTINE VALLASTER
PATRICK VELTE
CHRISTIAN VÖGTLIN
STEPHAN WAGNER
BARBARA E. WEISSENBERGER
ISABELL M. WELPE
HANNES WINNER
THOMAS WRONA
THOMAS ZWICK

Leon Specht, An Empirical Analysis of European Credit Default Swap Spread Dynamics 1

Moritz Setzer, From investor to entrepreneur – An explorative study of the entrepreneurial behaviour of investor-entrepreneurs 43

Michael Keck, Die Bedeutung der Unternehmenssteuerung für die Sicherstellung einer wahrheitsgemäßen Berichterstattung: Eine kritische Analyse 68

Niklas Behne, Going-Concern-Modifizierung des Bestätigungsvermerks – Eine kritische Analyse der Einflussfaktoren und Konsequenzen 96

Fabio Krüger, Impact of Team Agility on Team Effectiveness: The Role of Shared Mental Models, Team Empowerment, and Team Reflexivity 123

Julian Schneider, Förderung der Transparenz oder Kosten für die Aktionäre? Beurteilung der Regulierung von Proxy Advisors anhand der U.S.-Marktreaktion 148

Anna Bogner, Memory asymmetries in experiential and material purchases: The role of self-expression 163

Carina Menhofer, Kritische Analyse der umsatzsteuerlichen Behandlung von Gutscheinen seit dem Jahressteuergesetz 2018 188

Jonas Pütter, Is Cash (the only) King? – Eine kritische Analyse der Aussagekraft von Kapitalflussrechnungen nach IAS 7 219

Markus Klepsch, Do birds of a feather always flock together? A multidimensional examination of homophily in crowdfunding 237

Published by Junior Management Science e.V.

Junior Management Science 8(1) (2023) 1-42

Junior Management Science

journal homepage: www.jums.academy

An Empirical Analysis of European Credit Default Swap Spread Dynamics

Leon Specht

Leibniz Universität Hannover

Abstract

I analyze the dynamics of European credit default swap spreads by estimating CDS spreads via an extension of the structural credit risk models by Black and Cox (1976) as well as Leland (1994), the so called CreditGrades model proposed by Finger et al. (2002). Using two different procedures in approximating the asset volatility surface of obligors, the models are calibrated by means of historical equity volatility and volatility extracted out of at-the-money options. I discover that model performance strongly depends on the distribution of input parameters clustered by economical sectors. Model spreads exhibit significant correlation with market spreads and seem to predict market spreads contingent on sectors and model calibration techniques. The gap between model and market spreads, derived model spreads and empirical market spreads are analyzed by running panel regressions in fashion of Collin-Dufresne and Goldstein (2001) and Bedendo, Cathcart, and El-Jahel (2011). These show that times of disconnectedness between credit and equity markets, model inherent misspecifications as well as possible market inefficiencies can contribute to the inability to estimate spreads reliably. Robustness checks show that determinants of gap, model and market spreads are sector specific, time varying and tenor dependent.

Keywords: Credit Risk; Credit Risk Modelling; Structural Models; Credit Risk Management; Quantitative Finance.

1. Introduction

Credit risk is among the most fundamental economical risks. It emerges from the act of lending and borrowing capital as one of the main objectives of the financial system and allows for an efficient allocation of capital from surplus sectors to sectors in demand for capital, increasing economical welfare. In this context, the financial system fulfills two major functions: a reduction in information and transaction costs as well as the enhancement of trading, diversification and risk management. Since financial intermediation and their institutions facilitates the allocation of capital within an economy mainly due to asymmetric information and accompanying comparative advantages in associated monitoring costs, banks have typically the largest exposure to credit risk (Allen & Santomero, 1997; Diamond, 1984). Due to this fact, banks and other financial institutions are in constant need to quantitatively manage risk via mathematical models, thus limit the downside of lending not only by internal risk thresholds, but also regulatory requirements, such as Basel III and accounting loss allowance as well as risk appetite rules. With the rise of credit derivatives in the 1990s, the need for assessing the risk and deriving fair value pricing emerged. Financial institutions came up with quantitative models to control risk in new markets. One of those models is the CreditGrades model that was developed in the early 2000's by investment banks to establish an industry standard open source quantitative credit risk model for publicly traded firms. Besides the commercial use of those models, the academic world became more and more interested in their performance. Up to the year 2010, a lot of researchers considered commercial credit risk models in academic settings. Since then, commercial credit risk models decreased in popularity because the market of credit derivatives matured over time, so that financial institutions relied more on well tested and experienced models. Here my thesis steps in and tries to shed light onto the CreditGrades model and its performance, as it was primarily tested only in the North American derivatives market in the early 2000's with short observation windows. The aim of my study is to estimate the CreditGrades model for a sample of European Credit Default Swap (CDS) contracts with long observation windows, covering multiple crisis and tranquil periods in order to assess its performance in various economical conditions. Furthermore I employ different model calibration procedures using information from European equity as well as option markets to test the hypothesis if the model performs better with forward-looking information extracted from option markets as often cited in the literature covering

DOI: https://doi.org/10.5282/jums/v8i1pp1-42

the North American market. After the estimation and performance evaluation, I proceed with analyzing the gap between model and market spreads. To do so, I employ panel regressions in order to identify drivers of divergence and shedding light onto possible model misspecifications. Furthermore I try to identify drivers of model and market spreads in the spirit of Collin-Dufresne and Goldstein (2001), who consider credit spreads on corporate bonds. Related to that, I test if the CG Model entails predictive behavior, which is exploited by capital structure arbitrage trading strategies. Additionally I employ time period panel regressions to reveal whether the determinants are time invariant or if certain factors are more influential in specific economical environments.

To the best of my knowledge I provide the largest study that tests the performance of the CreditGrades model in its respective specifications and determinants of credit default swap spreads considering the European credit market. Furthermore I am the first who considers the performance of the CreditGrades model in the Corona crisis of 2020 and its determinants of the gap, empirical CDS spreads and model spreads both on a European and international level.

My thesis is structured as follows. Section 2 deals with credit risk and associated derivatives. Then, section 3 provides an overview of the importance of structural credit risk models and its implication for identifying determinants of CDS spreads. Afterwards, section 4 outlines the methodology of my study, including my model choice, the model calibration procedures as well as the subsequent correlation and pricing analysis and the multi-factor model CDS determinants estimation procedure. Section 5 introduces the data used in my thesis. My empirical analysis is located in section 6 and starts with descriptive statistics on my sample composition involving respective summary statistics, followed by the results of my correlation and pricing analysis, showing that model performance strongly depends on the distribution of input parameters clustered by economical sectors. These results are then analyzed in more detail by means of panel regressions covered in subsections 6.3 to 6.6. I proceed with robustness checks in chapter 7 where I control my results for different tenors and time periods, finding that determinants of gap, model and market spreads are sector specific, time varying and tenor dependent. Finally section 8 concludes.

2. Credit Risk and associated Credit Derivatives

Credit risk describes the possibility that the borrower cannot meet his/her financial obligations, thus defaults on the capital amount outstanding. Credit risk is hence not driven in an aggregate, but consists of individual factors and drivers which are called mathematical components of credit risk in the following analysis. The Bank for International Settlements (BIS) depicts these components as credit risk parameters in their principles for the management of credit risk (BIS (2000)) which coincide with Basel regulation as follows. Probability of Default (PD): the statistical likelihood a default occurring within the lifetime of the loan contract. The default is defined by a default event, which is typically measured in delinquency days or days past due of 90 days. Exposure at Default (EAD): the amount of capital outstanding and thus at risk. Loss given Default (LGD): the actual loss contingent on the default of the obligor measured in percentage points of the exposure. LGD is typically smaller than 1, since the proceeds from repossessing, required collateral as well as other guarantees can mitigate the loss risk and thus the LGD largely depends upon the type of default and the subsequent settlement. The inverse of the LGD is called Recovery Rate (RR) and is defined as 1 - LGD. The resulting credit risk is then a function of all mathematical components

$$CreditRisk = f(PD, EAD, LGD).$$

One could assume that credit risk has to be accepted once a loan is granted, thus if a loan or bond would be issued, the risk of failing repayment could not be mitigated ex-post and hence has to be compensated by either high interest rates or high coupons ex-ante. The role of risk management would be constrained to the instruments of limited exposure, ex-ante time invariant low PDs and high RRs. This situation was largely present up to the 1990s, since at that time, financial innovation led to the emergence of so called credit derivatives (Brandon & Fernandez, 2005). These instruments allow the financial market to allocate and distribute pure isolated credit risk to those parties, who have risk bearing capacities and to those who would like to reduce their credit risk appetite. This aspect marks a fundamental difference to tradeable debt of entities with corresponding credit risk exposure (e.g. corporate bonds), since credit derivatives enable the pure/isolated credit risk to be tradeable as they eliminate the necessity to hold the asset subject to credit risk itself. Hull (2012) classifies credit derivatives in two main categories, which are the following: single-name and multi-name credit derivatives. Since my study considers the dynamics of CDS spreads, I will not give a detailed overview of multi-name credit derivatives such as Collateralized Debt Obligation (CDO) or basket CDS, but rather focus on the single-name branch of credit derivatives. The most common single-name instrument is the CDS, which was first introduced by J.P. Morgan in 1997 (Chan, Guo, Lee, & Li, 2019). CDS offer protection against the default of corporations or sovereigns. Corporations are called reference entities upon which the CDS is issued. These reference entities are not directly part of the derivatives contract, since they are bilateral negotiated between the issuer of the CDS (most often investment banks or insurance companies) and the buyer of the CDS (most often hedge funds or other banks). This setup allows for credit risk to be discretely traded and shifted across financial market participants. The typical CDS contract consists of two sides, the buyer and the seller side. The buyer side obtains credit protection for the issued reference entity in case of default by paying period premiums to the seller of the CDS, the so called CDS spread, which is measured in Basis Points (bps) of the notional amount. The seller of the CDS, however, has to settle the contract in case of an occurring credit event either physically or in cash. Thus, the CDS buyer is entitled to sell the underlying bonds at their notional amount,

whereas the CDS seller has to buy the bonds at the notional amount. These payment streams boil down to exchanging streams of cashflows wherefore this product is considered as a swap agreement. The face value of the debt that the CDS buyer is entitled to sell is called the notional principal. The debt itself is called the reference obligation (Hull (2012)).

Moreover, CDS products are defined by their maturity or tenor for which the protection will last. A typical maturity for a CDS is five years. Thus if the reference entity would experience a credit event within five years, the buyer of the CDS would be entitled to receive credit protection by the mechanism described above. Another crucial aspect is the definition of a default event. The International Swaps and Derivatives Association (ISDA) defines three main types of credit events: filing for bankruptcy, defaulting on payment and the restructuring of debt. This differentiation has material impact on CDS valuation in markets, since they carry different degrees of default risk profiles dependent on the inclusion of specific credit events. Berndt, Jarrow, and Kang (2007) find that the premium for exposure to restructuring risk amounts on average between 6% and 8% of the value of protection against non-restructuring default events.

As any derivative, credit derivatives fulfill the three central functions of hedging, arbitrage and speculation. The function of hedging considers CDS as an insurance type product for mitigating the exposure to credit risk by paying periodic premiums with the aim to transfer the risk. In a more precise manner, this allows for instance to create positions of risk-free corporate bonds since the investor can buy the bond with its time to maturity and match this by buying a CDS written on the same bond issuer with the corresponding tenor. The aspect of arbitrage covers a specific strategy called capital structure arbitrage. Capital structure arbitrage is based on the idea that there exists mispricing between equity and credit markets and the subsequent exploiting of this divergence from an equilibrium state by applying trading strategies (Wojtowicz, 2014). This type of arbitrage was very popular during the early 2000s, since at that time hedge funds started to explore credit derivatives and their applications in trading strategies. In accordance to the attention in the practitioners world, the academic sphere also took capital structure arbitrage under closer consideration. The most influential study was performed by Yu (2006) who have considered the risk and return profile of capital structure arbitrage by calibrating a structural credit risk model, the CreditGrades Model (CG Model), to a sample of 261 North American industrial obligors between 2001 and 2004. He implements trades by selling CDS and buying equity when the model predicts that CDS are overpriced and buying CDS and shorting equity when the model predicts that CDS are undervalued. He assumes that the trades eventually convergence when the model spread is equal to the market spread. Yu (2006) finds that individual trades are very risky resulting in complete drawdowns of capital. The convergence rate of trades are found to be as low as 10% and annualized trading returns for a 30 day holding period are negative or close to zero. If a 180 day holding period is implemented, the mean return is at its maximum 2.78%. Yu (2006) tests his results for statistical arbitrage using the procedure described in Hogan, Jarrow, Teo, and Warachka (2004) and has to conclude that in his setting, there are no significant arbitrage returns existent. The work of Yu (2006) was afterwards extended by the study of Bajlum and Larsen (2008) who use an advanced version of the CG Model calibrated to one month at-the-money (ATM) put option implied volatilities. Cao, Yu, and Zhong (2011) later support the findingy of Bajlum and Larsen (2008). Bajlum and Larsen (2008) achieve higher and statistical significant trading returns, in case for speculative grade obligors up to 4.61%. Huang and Luo (2016) use call option implied volatilities and derive a persistent trading strategy that produces significant risk adjusted excess returns, hence supporting the finding of Bajlum and Larsen (2008). Duarte, Longstaff, and Yu (2007) take a different approach in testing an array of different trading strategies, including capital structure arbitrage. They use the CG Model and are able to create positive and significant excess returns after trading costs. Imbierowicz and Cserna (2008) calculate CDS spreads using the CreditGrades model, the structural credit risk model by Leland and Toft (1996) as well as the model by Zhou (2001) and test capital structure arbitrage between 2002 and 2006 on 808 obligors and can support the findings of Duarte et al. (2007). As with any other structured financial product, speculation can be one field of application. If a speculator concludes that the true PD is higher than the market implied PD, he can buy default protection at a discount that pays a positive yield in case of default. On the other hand, if the speculator believes that the true PD is lower than the market implied PD, he can sell credit protection at a premium and collect the overvalued spread thereafter.

Regarding the valuation of CDS spreads, there are many theories, models and methodologies. However there exists one quantitative approach based on a fair value pricing mechanism widely regarded as the main pricing methodology of individual plain vanilla CDS, called the Hull-White Model. Hull and White (2000) base their approach of deriving CDS spreads on the present value of the so called fixed and floating/contingent leg of the CDS cashflow stream. The fixed leg represents the periodic spread payments transferred from the buyer to the seller. The floating/contingent leg represents the seller side, since the seller has to buy the underlying bonds in case of default. Thus, the cashflow transferred from the seller to the buyer is contingent on the reference entity experiencing a credit event. The main idea of the model is that the present value of all premiums payed should equal the present value of the expected payoff of the CDS for both participants, resembling no arbitrage conditions. Hull and White (2000) assume that default events, the risk free rate and the recovery rate are mutually independent from each other and that the claim in case of default is defined to be the notional amount additional to accrued interest. The risk neutral probability of no credit event happening during the lifetime of the swap is one minus the probability that a credit event will occur up to

time T,

$$\pi = 1 - \int_0^T q(t)dt, \tag{1}$$

with $q(t)$ the risk neutral default probability density at time t. The expected present value of the fixed leg is determined by the payments until a credit event happens or until the maturity of the contract at time T in case no credit event occurs. If the default occurs at $t < T$, the Present Value (PV) of the payments is $w[u(t) + e(t)]$, with w the total payments per year made by the CDS buyer, $u(t)$ the PV of payments at the rate of \$ 1 per year on payment dates between $t = 0$ to t and $e(t)$ the PV of an accrual payment at time $t - t*$ with $t*$ as the preceding payment date after t. Hence the expected PV of all payments is given by

$$w \int_0^T q(t)[u(t) + e(t)]dt + w\pi u(T). \tag{2}$$

The market value of the reference obligation just after the reference entity has defaulted is the recovery rate multiplied with the sum of the notional amount plus accrued interest

$$P - RP[1 + A(t)] = P[1 - R - A(t)], \tag{3}$$

with P as the notional principle, R the recovery rate and $A(t)$ as the accrued interest up to time t measured in percentage of the notional amount. If the above described assumptions apply, the risk neutral expected payoff can be written as

$$1 - [1 + A(t)]\hat{R} = 1 - \hat{R} - A(t)\hat{R}, \tag{4}$$

with $\hat{R}$ as the expected recovery rate in the risk-neutral world. Then the present value of this expression indicating the expected payoff is equation (4) integrated over the lifetime of the CDS

$$\int_0^T [1 - \hat{R} - A(t)\hat{R}]q(t)v(t)dt, \tag{5}$$

with $v(t)$ as the present value of \$ 1 received at time t. The present value of the CDS to the protection buyer is the present value of the expected payoff in case of default (equation (5)) after subtraction of the present value of the periodic payments made by the buyer in equation (2)

$$\begin{aligned} &\int_0^T [1 - \hat{R} - A(t)\hat{R}]q(t)v(t)dt \\ &- w \int_0^T q(t)[u(t) + e(t)]dt + w\pi u(T). \end{aligned} \tag{6}$$

Hull and White (2000) conclude that the CDS spread s is the value of w that is setting equation (6) to zero

$$s = \frac{\int_0^T [1 - \hat{R} - A(t)\hat{R}]q(t)v(t)dt}{\int_0^T q(t)[u(t) + e(t)]dt + w\pi u(T)}, \tag{7}$$

such that the spread is defined as the total amount of payments within a year measured as an annualized percentage of the notional amount of the CDS reference entity.

3. Literature Overview

Chapter 3 gives an overview of the most prominent credit risk models cited in the academic literature. Since my thesis considers the use of an extension of a structural credit risk model, I proceed in outlining the main differences to so called reduced-form models and provide pricing implications for credit risk thereafter. Afterwards I shed light into the implications of these models providing guidance for the empirical determinants of CDS spreads cited in the literature so far, setting up the rational for the model choice in my empirical part closer assessed in chapter 4.

3.1. Structural Credit Risk Models and their extensions

In the academic literature, credit risk modeling can be split into two fundamental types of models. The first type, called the *structural credit risk model*, establishes a relationship between the capital structure of a firm and the corresponding default probability. The main intuition of these models is that corporate liabilities can be viewed as contingent claims on the firm assets (Lando, 2009). The second strain of credit risk models are called *reduced-form models*. They are based on the idea that credit risk is driven by exogenous variables, such as hazard intensity rates or credit migration and model it by a random stochastic Poisson processes that is independent of the capital structure of firms (Klieštik & Cúg, 2015). The main advantage of these types of models is that they do not require to obtain information on the capital structure of individual firms (Klieštik & Cúg, 2015). However, this is also the the biggest drawback of reduced-form models since they do not allow for interpretation using economic theory. These types of models deliver mathematical tractable solutions, but lack the depth and economical background needed to derive and identify economical drivers and determinants of default risk. For this reason, I will not cover these types of models in detail. The most prominent reduced-form models are formulated by Duffie and Singleton (1999); R. Jarrow and Protter (2004); R. A. Jarrow and Turnbull (1995) and Madan and Unal (2000).

The first and arguably most important structural model was developed by Merton (1974). He was the first who applied modern option pricing theory developed by Black and Scholes (1973) to credit risk modeling by establishing the methodology that a company's equity can be modeled as a call option written on the firm's assets. Thus model implied probabilities of default are endogenous to the capital structure of the firm and are not driven by exogenous shocks or hazard rates. Since the Merton model relies on option pricing theory, the so called Black-Scholes assumptions apply (see appendix A). The main idea of Merton (1974) is that the capital structure of a company consists out of equity and debt. Debt is modeled as a single zero-coupon bond outstanding D_t with face value K and maturity T. Equity is defined as the market value of equity E_t. The assets of a company are derived by the balance sheet identity equation $V(t) = E_t + D_t$. Merton adjusts the assumptions of Black and Scholes (1973)

that the value of equity follows a Geometric-Brownian Motion, since now the assets of a company follow a Geometric Brownian Motion[1]

$$dV(t) = rf_t dtV(t) + \sigma V(t)dW_t, \tag{8}$$

with rf_t as the continuously compounded risk-free interest rate determining the asset drift, σ the asset return volatility determining the firm asset diffusion and dW_t a Wiener process stochastically driving the asset value over time. If $v(t) \geq K$ then debtholders are payed in full and the residual value of $v(t)$ is $v(t)-K$ and thus is distributed among equity holders. In case of $v(t) < K$, the debtholders can only be partially payed and the equity holders receive nothing. The company has defaulted on its debt at maturity T. This yields the payoff profile of a European call option written on the firm assets $v(t)$ with strike K and maturity T

$$E_t = max(V(t)-K;0). \tag{9}$$

Now the Black-Scholes formula for a European call can be applied, giving

$$E_t = V(t)\Phi(d_1) - Ke^{-rf_t(T-t)}\Phi(d_2) \tag{10}$$

with

$$d_1 = \frac{ln(\frac{V(t)}{K}) + (rf_t + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}} \quad \text{and}$$
$$d_2 = \frac{ln(\frac{V(t)}{K}) + (rf_t - \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}. \tag{11}$$

The default event now describes the possibility that the European call of equity holders matures out-of-the-money, which yields the risk-neutral default probability[2]

$$PD_t = Prob(V(t) < K) = \Phi(-d_2). \tag{12}$$

A common problem addresses the calibration of the Merton model, since the value of assets $v(t)$ and the corresponding asset volatility are both unobservable parameters. Merton (1974) solves this by assuming that E_t is also driven by a Geometric Brownian Motion and by applying Ito's Lemma, the identity between observable equity volatility and unobservable asset volatility are given as follows

$$V(t)\sigma\frac{\partial E_t}{\partial V(t)} = E_t\sigma_e. \tag{13}$$

By inserting the call delta expression using Black-Scholes in (13), obtaining

$$E_t\sigma_e = \Phi(d_1)\sigma V(t). \tag{14}$$

[1]Because Black-Scholes derive their formula in a risk-neutral world under the risk-neutral measure Q, the asset drift is now determined by the risk free rate, thus μ is replaced with rf_t.

[2]Whereas the risk-neutral probability is always smaller than the physical probability of default, if and only if $rf_t < \mu$.

Equation (10) and equation (14) can both simultaneously solved for $v(t)$ and σ to derive probabilities of default by inserting the asset value and asset volatility in equation (12). The described model setup above relies on an number of simplifying assumptions that do not correspond to the real-world environment in which firms operate and default risk emerges. Thus various follow-up models address these shortcomings by implementing different aspects as follows.

Point of default. The Merton model assumes that a company can only default on its debt at maturity T. Black and Cox (1976) relax this assumption by implementing a so-called exogenous default barrier, active throughout the lifetime of the company. The barrier is derived by exploiting the fact that there exists a probability distribution with a corresponding Laplace transformation explicitly known for the Brownian Motion first hitting a certain level (the default barrier) assuming time invariant parameters for the asset drift and diffusion (Duffie & Singleton, 2003). If the firm assets fall under this barrier, the company is considered to have defaulted. Hence, this model is frequently referred to as first passage time model. The default barrier has to be only hit once to constitute a default event. In Collin-Dufresne and Goldstein (2001) this first passage model is extended by modeling a stochastic default barrier which is still exogenous but time-dependent. The amount of debt is conditional on the changes in firm value which results in a stochastic mean-reverting debt-to-equity ratio. Leland and Toft (1996) are the first who attempted to derive an endogenous structural credit risk model by considering optimal capital structures of a firm to maximize the market value of equity E_t in order to optimize the amount and the maturity of its debt.

Debt structure. To characterize the entire debt of firms as one zero-coupon bond outstanding rarely meets the actual debt structure of firms. Thus the structural model proposed by Geske (1977) addresses this aspect and models debt as a coupon bond, where each coupon is viewed as an option on the firm value. This enables the firm to default not only at maturity, but at any coupon date. Since the assumption that a stable debt structure is maintained throughout the lifetime of the firm holds, equity is issued to settle the coupon payments. The firm defaults, if equity holders do not issue new equity due to unfavorable investor conditions, resulting in a failing coupon payment defined as a credit event.

Stochastic processes for asset values and interest rates. The standard structural model by Merton (1974) is derived by using a Geometric Brownian Motion to drive the asset value process. A standard GBM cannot account for sudden jumps in the firm value by construction. Hence, a firm never defaults due to unexpected short term changes in its firm value. This aspect was empirically criticized by the assessment of model estimated short term credit spreads being too low, especially for low rated companies. Merton (1976) adjusted the standard GBM asset value process, introducing a jump diffusion process in form of a Poisson process. This type of model is often called hybrid credit risk model, since it combines the economical foundations of structural models and extends those by Poisson processes, also found in reduced-form models,

in order to obtain mathematical tractable results that better match the empirical data. Zhou (2001) applied the extension by Merton (1976) to model credit spreads of corporate bonds and concludes that by incorporating a compound Poisson jump diffusion process, the performance of the Merton model improves significantly. This improvement is especially achieved by accounting for different term-structures of credit spreads unique to obligors. The work of Boyarchenko (2000) as well as Hilberink and Rogers (2002) build on the work of Zhou (2001) by introducing Lévy jump processes. Besides the assumptions of the stochastic asset values process, the assumption of a constant interest rate does not seem realistic, since market rates are typically not time invariant. Kim, Ramaswamy, and Sundaresan (1993) are the first who considered stochastic interest rates explicitly by using the influential CIR short rate model proposed by Cox, Ingersoll Jr, and Ross (1985). Longstaff and Schwartz (1995) consider the same methodology but rely on a Gaussian one-factor model derived by Vasicek (1977) for the short rate.

3.2. Determinants of Credit Default Swap Spreads

The literature covering the empirical analysis of determinants of credit default swap spreads has its roots in the analysis of credit spreads. Credit spreads are defined as the difference in yield of corporate bonds and a risk-free benchmark yield, such as government bonds (Collin-Dufresne & Goldstein, 2001). Credit risk models propose that this difference emerges due to the exposure to credit risk in accordance to different credit risk factors consistent with theoretical models. However, empirical research assesses this assumption by testing other factors driving credit spreads. This field of research was initiated by the seminal work of Collin-Dufresne, Goldstein, and Martin (2001) who analyzed the determinants of changes in credit spreads for different credit rating groups, leverage groups and maturity groups. They base their analysis on variables from structural credit risk models, such as the risk-free rate, slope of the yield curve, leverage, volatility, the probability of a downward jump in firm value and business climate. They then regress changes in credit spreads from 1988 to 1997 on these proposed factors and conclude that the explanatory power of the regressions are in general rather low, explaining only 25% of the variation in credit spreads. Even if other control variables, such as financial and economic variables, are introduced, the results do not significantly change relative to the initial model setting. An afterwards conducted Principal Component Analysis (PCA) of the regression residuals shows that the model lacks a common risk factor, since the first component accounts for 76% of left over variation in credit spreads. Collin-Dufresne et al. (2001) conclude that local supply and demand shocks, independent between treasury and equity markets, can be one feasible explanation for the missing common aggregated systematic risk factor.

Tsuji (2005) was the first who called the findings of Collin-Dufresne et al. (2001) the *credit spread puzzle*. He conducts the same econometric analysis for Japanese bonds considering determinants implied by theoretical models, illiquidity, risk aversion and expectations, credit ratings, business cycle and other bond characteristics. Despite this extensive approach, he mostly confirms the puzzle. Campbell and Taksler (2003) explicitly consider the effect of firm-level volatility, but employ regressions to explain levels of credit spreads. They conclude that the idiosyncratic equity volatility can explain variation in credit spreads comparable to credit ratings. Thus, idiosyncratic risk has significant explanatory power (Campbell & Taksler, 2003). Cremers, Driessen, Maenhout, and Weinbaum (2008) support this finding as they consider option-implied volatilities as the level of implied volatility and the implied volatility skew seem to matter for the explanation of credit spreads.

In the subsequent years after the credit spread puzzle was discovered, the focus of the academic literature considering determinants of credit risk shifted from analysing credit spreads to investigating credit default swap spreads. As outlined in chapter 2, credit derivatives have some advantages regarding investors/speculators/arbitrageurs when comparing them with their underlying bonds. But also for researchers CDS entail advantages when assessing them in econometric settings. While CDS spreads are comparable to credit spreads of corporate bonds, they do not require to specify a risk-free yield curve since the observable price for a CDS is already expressed as the spread itself. Thus by choosing CDS spreads as the dependent variable, possible effects arising from the choice of the risk-free rate can be mitigated. Houweling and Vorst (2005) show that the choice of a risk-free rate has substantial impact on the estimation of credit risk premia. In their study only the usage of swap or repo rates results in reliable estimation of credit risk premia. Ericsson, Jacobs, and Oviedo (2009) also mentions the aspect of accounting for coupon payments and possible effects triggered by choosing the corresponding framework to remove coupon payments. Another favorable argument for declaring CDS spreads as the subject of analysis are liquidity reasons. Due to the fact that derivatives are in general more frequently traded, higher observation frequencies (e.g. daily) with a better data quality favor CDS spreads. In addition, Blanco, Brennan, and Marsh (2003) show by assessing the rate of spread change in both markets conditioned on a change in the credit rating, that a change in credit quality is faster incorporated in the credit derivative markets than in the corresponding corporate bond market. This hints at a more efficient incorporation of new information in derivative markets when considering the likelihood of a credit event. Additionally, Di Cesare and Guazzarotti (2010) state that CDS are inherently more standardized than bonds due to missing aspects like coupon structure, callability, convertability and other bond specific factors. Hence, increasing the level of comparability and standardization within a study design improves the implementation and interpretation of subsequent empirical results.

The first study considering determinants of CDS spreads was performed by Aunon-Nerin, Cossin, Hricko, and Huang (2002). Their study was motivated by the work of Collin-Dufresne and Goldstein (2001) published in the previous

year. By considering so called "structural variables" motivated by credit risk pricing theories and applying the same methodology as Collin-Dufresne et al. (2001), Aunon-Nerin et al. (2002) find significant explanatory power reaching 82% in R^2 for the cross-sectional variation in level CDS spreads. Furthermore, variables used as proxies for equity market information (e.g. stock prices and its variance) as well as stochastic interest rates and credit ratings deliver robust results when pricing credit risk in the cross-section. The work of Skinner and Townend (2002) was the first paper officially published in an academic journal that empirically examines CDS spreads. The study contains 29 US dollar denominated CDS spread time series observed between 1997 to 1999 and thus also captures the Asian Currency Crisis in 1997. Skinner and Townend (2002) apply the same methodology as Aunon-Nerin et al. (2002) by analyzing CDS spreads in an OLS regression framework. They propose five factors motivated by option pricing models and test those in their subsequent regression analysis. They find that four factors, the risk-free rate, interest rate volatility, time to maturity and the yield on the underlying bonds are statistically significant, whereas the exercise price, defined as the payable amount by the CDS seller in case of default, does not appear to be significant. Ericsson, Reneby, and Wang (2006) use a set of structural models to predict CDS spreads and afterwards regress the residuals on default and non-default proxies, concluding that there are little or no significant factors driving the residuals. In consequence, credit risk models seem to adequately price the credit risk in credit derivatives markets.

Besides structural models that identify structural variables as determinants of CDS spreads, some researchers also considered volatility and jump risk as factors that contribute to the explanatory value of the dynamics in credit derivatives markets. These studies are mainly motivated by the findings of Campbell, Lettau, Malkiel, and Xu (2001), Campbell and Taksler (2003) and Goyal and Santa-Clara (2003). Benkert (2004) was the first to consider the effect of equity volatility on default risk in CDS markets. Benkert (2004) follows Campbell and Taksler (2003) in relying on fixed effects panel regression analysis that control for time and firm effects. His final sample includes spreads for 120 reference entities between January 1999 and May 2000, totaling to 26.478 quotes. In his regression model, he controls for liquidity, interest coverage, firm leverage, profitability, credit rating, annualized historical equity volatility over 180 calendar days and Black-Scholes option implied volatility derived from at-the-money stock options. In their setting, option implied volatilities show high statistical significance, contributing 5 to 10 percentage points to the R^2 in reference to the setting where no volatility control variable is employed. Furthermore, Benkert (2004) confirm the results of Campbell and Taksler (2003) that historical volatilities are also significant and have the expected sign. In comparison to implied volatility, however, the associated R^2 is significantly smaller, indicating higher efficacy of implied volatility in the explanatory value of CDS premia. The importance of jump risk was considered by Zhang, Zhou, and Zhu (2009) who employ OLS regressions and confirm the results of Benkert (2004). In their study, volatility risk accounts for 48% of the variation in CDS spread levels, whereas jump risk alone accounts for 19%. Hence volatility and jump risk are both the most important factors driving CDS spreads. Zhang et al. (2009) also use factors such as credit ratings, macroeconomic factors and firm specific balance sheet information and are able to explain up to 73% of CDS spread levels, both valid for investment grade and sub-investment grade reference entities.

A number of empirical studies assess the role of liquidity in credit derivative markets, mainly motivated by the seminal work of Amihud (2002) and Pástor and Stambaugh (2003) who identified liquidity risk as an important risk factor in the context of equity asset pricing. Besides illiquidity in equity markets, debt markets also contain priced illiquidity risk factors as shown by De Jong and Driessen (2012) and Bongaerts, De Jong, and Driessen (2017). Regarding CDS markets, Tang and Yan (2007) are the first who examined the role of liquidity in credit derivatives markets. The CDS data used in their study cover the period from June 1997 to March 2006, including 27 industries of U.S. corporate senior unsecured CDS, denominated in US Dollar with a maturity of 5 years. They perform panel regressions with robust standard errors and monthly time dummies to control for issuer clustering in time series correlation's. As a result, they find that liquidity effects arsing from liquidity characteristics account for 13.2 bps in the cross-section, whereas in a beta pricing framework, the liquidity risk premium is estimated to be about 10.9 bps (Tang & Yan, 2007). In later studies performed by Chen, Fabozzi, and Sverdlove (2010), Bongaerts, De Jong, and Driessen (2011) and Junge and Trolle (2015), the role of liquidity in CDS markets was further analyzed, providing additional evidence on the fact that liquidity risk has an influence on CDS valuation.

Another obvious CDS determinant that can be controlled for is the credit rating of the obligor. Hull, Predescu, and White (2004) consider the effect of rating announcements on changes in CDS spreads before and after the change in credit rating. They obtain 233.620 individual CDS spread observations between 1998 and 2002 and conclude that the review for a downgrade in credit status contains significant information, whereas counter intuitively the downgrade itself and negative outlooks do not. Positive rating changes do not exhibit the same significance in the CDS dynamics compared to negative reviews. Given these findings, Hull et al. (2004) subsume that the CDS market anticipates negative credit reviews before the actual downgrade happens. Thus, CDS spreads entail predictive characteristics when the likelihood of a downgrade is considered.

Besides firm-specific factors like volatility, illiquidity and the usage of structural variables indicating credit risk, also macroeconomic factors play an important role in the determination of factors driving CDS spreads. General market conditions and the overall state of the economy can provide meaningful information about the environment in which firms operate and default risk emerges. Due to this fact, a number of studies exist that consider the explicit explanatory power of

macro factors such as inflation, GDP growth, unemployment rate, business climate, consumer confidence and overall market sentiment. The role of macroeconomic factors are predominantly considered in studies that analyze credit spreads, most prominently in Collin-Dufresne and Goldstein (2001).

The above mentioned determinants need not be time invariant. In consequence, several studies examined so-called regime dependent determinants of CDS spreads. They typically define specific time periods, such as crisis periods, in which specific factors contribute more or less to the explanation in CDS dynamics than others. Alexander and Kaeck (2008) most prominently assess time dependencies in delta spreads of iTraxx Europe CDS indices by employing a Markov regime switching regression model. They find that spread changes exhibit severe regime dependent as well as sector specific behavior. In times of high turbulence, spreads are very sensitive to stock volatility, whereas in times of tranquility, spread changes are more sensitive to stock returns. In the volatile regime, the size of regression coefficients are significantly higher in comparison of the tranquil regime, contributing to previously unexplained parts of variation in credit spreads, e.g. in Collin-Dufresne and Goldstein (2001). In line with Benkert (2004), they conclude that structural variables of credit risk can contribute to the explanatory power of the model.

4. Methodology and Model Choice

The following section outlines the methodology used throughout my study. Section 4.1 starts with the selection and introduction of a structural credit risk model that is used as the reference point for the estimation of CDS spreads with structural input variables. Section 4.2 describes the different model calibration procedures employed to fit the model to market observables. Then section 4.3 assesses the performance of my fitted CG model by means of a conditional correlation analysis and popular forecasting measures. Section 4.4 sheds light on my selected econometric procedures employed to identify drivers and determinants of credit default swaps spread dynamics.

4.1. The CreditGrades Model

As outlined in section 3.1, structural credit risk models have the advantage of economical interpretability and allow for identification of the drivers and determinants of credit risk justified by economic theory. As my study aims to identify drivers of European CDS spread dynamics, I rely on structural credit risk models. Furthermore, I consider the availability of data and overall tractability of model implementation, given the time frame and overall setting of this thesis. When looking into structural credit risk models used to identify drivers of credit default swap spreads in more detail, one model is frequently cited and implemented by researchers. The so-called CreditGrades model (CG Model hereafter), was developed in the early 2000's by the RiskMetrics Group (now MSCI) in cooperation with the three major investment banks JPMorgan, Goldman Sachs and Deutsche Bank, to establish an industry standard open source quantitative single-name credit risk model for public traded firms. The CG model has to be seen as a complimentary addition to other commercial structural credit risk models that provide the same tractability and performance like the CreditMetrics model introduced in 1997 by JPMorgan. The main difference to other commercial structural credit risk models is that the CG model derived credit risk is determined on the base of CDS spreads and not probabilities of default. Thus, the model was among the first to address the need for a framework to capture credit risk in the credit derivatives market in the early 2000s. It was also motivated by increased regulation and capital requirements in that time. In addition to that, 2001 was the year of the burst of the dotcom bubble with resulting high default rates and the prominent default of Enron, further increasing the need for a framework of credit risk management open to any market participant (Thomas, 2002).

The authors describe the model setup and its implementation in the technical document by Finger et al. (2002), outlining that the CG model belongs to the group of structural credit risk models that link credit with equity markets. The model can be seen as an extension to the first passage time models of Black and Cox (1976) as well as Leland (1994), by modeling a default event as the market value of firm assets falling under a defined default barrier. The asset value process progresses as a geometric Brownian motion, however, Finger et al. (2002) assume that the asset drift term is zero, thus the firm is assumed to maintain a constant leverage ratio over time. The asset value process is defined as follows:

$$dV(t) = \sigma V(t) dW_t. \tag{15}$$

The authors note that for pricing credit risk through first passage time models, not the asset drift itself is relevant, but the asset drift relative to the default barrier. The constant leverage is obtained by either issuing more debt or pay dividends such that the drift of the debt and the drift of equity is the same. By no-arbitrage conditions, the asset drift should then also have the same dynamics as the debt and equity drift. This allows for the conclusion, that the "asset drift relative to the default barrier is zero."[3] The CG model defines the default barrier as the amount of firms assets that remain in case of default, which is the recovery value that debtholders receive. It is defined as $L \cdot D$, with L beeing the average recovery rate on debt and D the firms debt-per-share (Finger et al., 2002).

Another aspect that differs from standard structural models is that the CG model assumes a random default barrier. This stochastic threshold addresses the main problem of estimating artificially low short-term spreads when using classical Merton style default models, because asset processes that start above the barrier cannot cross the default threshold by pure diffusion in very short time intervals. This circumstance is most prominently documented in

[3] Finger et al. (2002), p.11.

Jones, Mason, and Rosenfeld (1984) and Eom, Helwege, and Huang (2004), who conduct extensive empirical work on comparing the performance of different structural credit risk models regarding their ability to estimate credit spreads for different maturities. Even though, Anderson and Sundaresan (2000) as well as Ericsson and Reneby (2001) use various extensions to the original Merton model, they still conclude a systemic underestimation of short-term credit spreads. The methodology of the CG model to introduce randomness to the average recovery rate L prevents the authors from using complex jump diffusion asset value processes or other advanced adjustments, but yields the same tractable estimations for short term spreads. Finger et al. (2002) justify the introduction of uncertainty to recovery rates based on empirical studies of recovery rates like the one of Hu and Lawrence (2000). Hu and Lawrence (2000) find large variability in recovery rates and contingencies depending on whether the default is due to financial and operational reasons or if the firm will be liquidated or restructured. Based on that, Finger et al. (2002) assumes that the recovery rate L is log normally distributed with mean $\bar{L}$ and corresponding standard deviation λ

$$\bar{L} = \mathbb{E}[L] \quad \text{and} \quad \lambda^2 = Var(log(L)). \tag{16}$$

The recovery rate is then denoted by

$$L = \bar{L}e^{\lambda Z - \lambda^2/2} \tag{17}$$

and the recovery value is given by equation (17) multiplied with the firms debt-per-share

$$LD = \bar{L}De^{\lambda Z - \lambda^2/2}, \tag{18}$$

where λ and $L \in \mathbb{R}^+$ and Z is a standard normally distributed random variable that is independent from the Brownian motion and equation (18) describes the stochastic default barrier. Z is considered to be unknown at $t = 0$, revealed only in case of default. Hence, Z can capture uncertainty in the level of debt-per-share and in its default barrier. Finger et al. (2002) conclude that there is some true time-invariant value of L that is not directly observable. This uncertainty in recovery rate allows that the default barrier can be hit suddenly, thus, implementing the same effect as a jump type diffusion asset value process (Finger et al., 2002). Figure 1 illustrates the described asset process and the stochastic default barrier in a condensed manner.

In summary, the firm will not default for the initial asset value V_0, if the following condition holds $\forall t$

$$V_0 e^{\sigma W_t - \sigma^2 t/2} > D\bar{L}e^{\lambda Z - \lambda^2/2}. \tag{19}$$

Finger et al. (2002) proceed in deriving the survival probability, i.e. the probability that the asset value in equation (15) will not hit the barrier defined in equation (18) until time t by introducing a process X

$$X_t = \sigma W_t - \lambda Z - \frac{\sigma^2 t}{2} - \frac{\lambda^2}{2} \tag{20}$$

to be able to transform equation (19) to

$$X_t = log(\bar{L}D/V_0) - \lambda^2 \tag{21}$$

which is normally distributed with

$$\mathbb{E}[X_t] = -\frac{\sigma^2}{2}(t + \lambda^2/\sigma^2) \quad \text{and} \quad Var(X_t) = \sigma^2(t + \lambda^2/\sigma^2). \tag{22}$$

Finger et al. (2002) then approximate the process X by a Brownian motion $\hat{X}$ with drift $-\sigma^2/2$ and diffusion σ^2. This step exchanges the uncertainty in the default barrier by the uncertainty in the asset value level at $t = 0$ to implement the methodology that the distance-to-default (DD)dd, i.e. the distance between asset value and default barrier, drives the credit risk in the model setup. The authors note that for $t \geq 0$, the moments of the original process X coincide with the moments of $\hat{X}$ and hence provide a valid approximation. In the next step, the distribution for first hitting time Brownian motions in the process form of $Y_t = at + bW_t$[4] is applied (e.g. in Musiela and Rutkowski (1998))

$$Prob(Y_s > y, \forall s < t) = \Phi\left(\frac{at - y}{b\sqrt{t}}\right) - e^{2ay/b^2}\Phi\left(\frac{at + y}{b\sqrt{t}}\right) \tag{23}$$

by inserting the drift term of $\bar{X}$ into $a = -\sigma^2/2$, setting equation (21) as y and shifting the time from t to $t + \lambda^2/\sigma^2$. The closed-form formula for the survival probability up to time t is given by

$$P(t) = \Phi\left(-\frac{A_t}{2} + \frac{log(d)}{A_t}\right) - d \cdot \Phi\left(-\frac{A_t}{2} - \frac{log(d)}{A_t}\right), \tag{24}$$

with

$$d = \frac{V_0 e^{\lambda^2}}{\bar{L}D} \tag{25}$$

and

$$A_t^2 = \sigma^2 t + \lambda^2. \tag{26}$$

The survival probability expressed in equation (24) implies that there exists the possibility that the firm defaults prior to $t = 0$ in the interval $(-\Delta t, 0]$, which is implied by the assumption that the default barrier is log-normal distributed and that the uncertainty in the default barrier is shifted to the Brownian motion starting value (Cao et al., 2011). This aspect results in the artefact that the probability of default in $t = 0$ is not zero. Finger et al. (2002) provide an alternative estimation of the survival probability addressing this aspect by integrating over the default barrier distribution which yields a survival probability expressed by a cumulative bivariate normal distribution. The closed form solution of the

[4] With constant a and b.

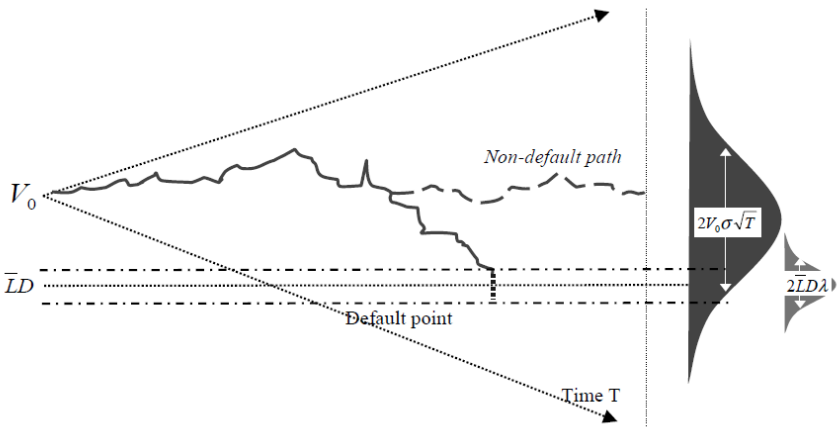

Figure 1: Illustration of CreditGrades Model, Finger et al. (2002).

exact survival probability was later derived by Kiesel and Veraart (2008). However, Finger et al. (2002) state that the numerical differences between the two approaches are small enough, such that for practical purposes, the initial implementation provides reasonable estimates.

To derive credit default swap spreads, the estimated survival probabilities must be transformed into a credit price. This is achieved by introducing two parameters, the risk-free rate r and the recovery rate of the underlying credit R. The important difference between $\bar{L}$ and R is that $\bar{L}$ describes the global average recovery rate across all debt classes and obligors and R depicts the individual recovery rate for the specific underlying asset subject to credit risk (e.g. the concrete bond issued by the obligor). A credit default swap is now priced by using the introduced methodology of Hull and White (2000) in chapter 2, requiring that the present value of the expected premium payments by the CDS buyer has to equal the present value of the expected payouts by the CDS seller in case of default.[5] Given the interest rate r and the survival probability derived in equation (24), the continuously compounded par credit default swap spread with maturity t estimated by the CG model equates to

$$s^* = r(1-R)\frac{1-P(0)+e^{r\xi}(G(t+\xi)-G(\xi)}{P(0)-P(t)e^{-r\xi}-e^{r\xi}(G(t+\xi)-G(\xi))}, \quad (27)$$

where $\xi = \lambda^2/\sigma^2$ and the G function is given by Rubinstein and Reiner (1991) as follows

$$G(t) = d^{z+1/2}\Phi\left(-\frac{log(d)}{\sigma\sqrt{t}} - z\sigma\sqrt{t}\right) + d^{-z+1/2}\Phi\left(-\frac{log(d)}{\sigma\sqrt{t}} + z\sigma\sqrt{t}\right) \quad (28)$$

and $z = \sqrt{\frac{1}{4} + \frac{2r}{\sigma^2}}$.

[5]For the complete derivation see appendix A of Finger et al. (2002).

The ability to price credit risk by the above described CG model is assessed by a hand full of researchers that predominantly consider the US CDS market. Byström (2006) was the first who considered the CG model for the European market. He uses the CG model to calculate CDS spreads on an aggregated level for eight iTraxx indices that cover the European area. All indices consist of 5 year maturity CDS contracts and are grouped into seven sectors (investment grade obligors) and one crossover index containing sub-investment grade obligors. He finds that model spreads are highly correlated with empirical spreads and constitutes that the CG model has predictive abilities since theoretical spreads tend to lead market spreads. This is mostly due to equity markets being faster in incorporating news than the CDS market (Byström (2006)). Furthermore he tests for autocorrelation and finds that market spreads exhibit severe autocorrelation, whereas model spreads do not. Byström (2006) states that this result provides reason for possible inefficiencies in the credit derivatives market, though equity markets seem to be more efficient. On this basis, Byström (2006) employs simple trading strategies that yield positive returns before transaction costs. The second most prominent study was performed by Yu (2006) as already outlined in chapter 2. In addition to that, Wagner (2008) assess CDS pricing via the CG model and compares it to the trinomial tree model proposed by Das and Sundaram (2001). They use a sample of 54 individual CDS extracted from the Markit CDX investment grade index covering the US market and the time period between May 2004 to July 2006. They find that both models' out-of-sample performance is poor with a low mean correlation in CDS spread deltas. However, when firm stock prices and CDS spreads highly correlate with the entire CDS market, the structural model fits better to empirical spreads and thus provides evidence on the fact that "model spreads fit better to market spreads given that less company specific credit risk is involved."[6]

[6]Wagner (2008), p. 26.

Bedendo et al. (2011) fit the CG model to 80 North American non-financial obligors between 2002 and December 2005 and analyse the resulting gap by running panel regressions for each sector on weekly deltas in the difference between model and market spreads. They choose to employ an advanced version of the CG model resting on the extraction of, not only option implied volatility, but also option implied estimates of asset volatility and the financial leverage of the firm using the implied volatility skew as the model input, taking option implied jump risk into consideration. Their main contribution is that the gap between model and empirical spreads tends to widen especially in crises times with high volatility. Especially the dynamics in equity implied volatility and volatility skew are major determinants of weekly gap deltas. Furthermore, liquidity factors and periods of a general disconnect between equity and credit markets result in high gaps. In line with Byström (2006) and Yu (2006), Bedendo et al. (2011) find that model spreads can act as predictors for changes in market spreads and contribute to the evidence that the CG model exhibits predictive abilities in credit markets.

Cao et al. (2011) employ two specifications of the CG model, one with option implied volatilities and the other with historical volatility of a 252-day horizon. Their sample consist out of 332 single name US credit default swaps observed between January 2007 until October 2009 and therefore also covering the global financial crisis in 2008. In order to assess the pricing performance of the CG model in respect to the different volatility estimations, Cao et al. (2011) conduct a pricing error analysis in which a pricing error ratio, calculated as the ratio of the implied volatility RMSE and the historical volatility RMSE, is analyzed. Afterwards they perform cross-sectional regressions with this ratio as the dependent variable in order to investigate if firm-level characteristics, such as spread volatility, credit rating, option volume and open interest, can provide some explanatory value. Their analysis shows that obligors with a high degree of spread volatility, option volume and lower credit ratings exhibit significantly lower pricing error ratios. This illustrates that implied volatility yields more credit market relevant information as when using historical volatility. They also tested the robustness of this finding by implementing different historical volatility estimators ranging from 22 to 1000 day horizons.

4.2. Model Calibration

In order to estimate CDS spreads using the CG model, model variables need to be calibrated to market observables. These observables cover the initial asset value of firms V_0, the asset volatility σ, the debt-per-share D, the global average recovery rate $\bar{L}$, the individual bond specific recovery rate R and the volatility of the default barrier λ.

Debt-per-share (DPS) is obtained by extracting liabilities from consolidated financial statements and dividing the liabilities by the number of shares outstanding. I follow Finger et al. (2002) and define liabilities as all debt issued that contributes in the financial leverage of a company. Hence all long- and short-term interest bearing financial obligations are included, whereas accounts payable do not enter my DPS calculation as they do not contribute to the financial leverage of the firm. Furthermore, I account for using consolidated financial statement data by adjusting the total debt for debt held by subsidiary companies. Since consolidated financial statements aggregate all debt, even though the subsidiary is not necessarily held a 100% by the parent company, I subtract the minority debt defined as $k \cdot MinorityInterest$ from the total debt, assuming that the debt-to-equity ratio k is equal to 1 and limiting the total amount of minority debt by not exceeding more than 50% of the total consolidated financial debt. The number of shares consist out of common shares outstanding plus preferred shares. As the number of preferred shares cannot be retrieved easily, I calculate them by retrieving the amount of preferred equity and divide it by the current equity price. Finally, debt-per-share is calculated by dividing the total debt, corrected for minority debt with the total amount of shares consisting of common and preferred shares. To avoid a look-ahead bias of the model, I update each parameter only when its information becomes available to the market. Hence, I keep the yearly consolidated financial data, such as total debt, minority interest and preferred equity, constant for the respective fiscal year. Common shares are retrieved quarterly, but are interpolated to obtain daily values since market participants can back out the number of common shares on a daily basis by dividing the current market capitalization by the current stock price. This way of estimating debt-per-share assures that the model provides an ex-ante view on the pricing ability of CDS spreads.

The instantaneous asset value V_0 and the asset volatility σ are both calibrated by assessing how the asset value behaves close to and far from the default barrier, relating the equity volatility to the asset volatility using Ito's Lemma:

$$\sigma_e = \sigma \frac{V}{S} \frac{\partial S}{\partial V}, \tag{29}$$

with the equity price of the firm denoted S and the corresponding equity volatility σ_e. The distance-to-default is given by η and is measured as the number of annualized standard deviations that separates the firms current equity value from the default barrier

$$\eta = \frac{1}{\sigma} log\left(\frac{V}{LD}\right) = \frac{V}{\sigma_e S} \frac{\partial S}{\partial V} log\left(\frac{V}{LD}\right). \tag{30}$$

The first boundary condition considers how V behaves close to the default barrier LD by assuming that the value of equity converges to zero when default becomes more probable. The value of the firms' assets then equal the contingent claim of debtholders in case of default defined as the default barrier itself

$$V|_{S=0} = LD \tag{31}$$

and close to the barrier adding the marginal rate of change in V based on a change in S

$$V \approx LD + \frac{\partial V}{\partial S} S. \tag{32}$$

The distance-to-default measure can now be expressed by substituting equation (32) into equation (30) in order to see how η behaves near the default barrier

$$\eta \approx \frac{1}{\sigma_e}. \tag{33}$$

The second boundary condition considers how V behaves far from the default barrier. Finger et al. (2002) assume that the asset value increases with the same speed as the equity value $S/V \to 1$. If this relation is substituted into equation (30), it follows that

$$\eta \simeq \frac{1}{\sigma_e} log\left(\frac{S}{LD}\right). \tag{34}$$

Finger et al. (2002) conclude that the most simple expression considering V and $e(t)aa$ that simultaneously satisfy the two boundary conditions is that $V = S + LD$. Hence, the value of assets consists of the value of equity plus the contingent claim of debtholders in times of default. The initial asset value V_0 is then given by

$$V_0 = S_0 + \bar{L}D, \tag{35}$$

with S_0 as the initial stock price and the corresponding volatility of assets given in equation (29)

$$\sigma = \sigma_e\left(\frac{S}{S + \bar{L}D}\right). \tag{36}$$

The estimation of σ is performed by a local approximation of the volatility surface, scaling the equity volatility σ_e by a gearing-ratio that accounts for leverage in the assets relative to equity only (Cao et al., 2011). The estimation of equity volatility is dependent on the research design. I employ two estimation procedures for equity volatility in order to be able to compare the pricing implications of using information from equity markets with forward-looking information extracted from derivative markets. The first procedure relies on historical volatility calculated from a 1000-trading-day moving average of daily equity returns that are afterwards scaled by $\sqrt{252}$ to obtain yearly volatility estimates. The choice of the length of the MA window is motivated by the studies of Duarte et al. (2007), Finger et al. (2002) and Yu (2006). For the other approach I obtain equity option implied volatilities derived from put and call options of the specific reference entities trading at 100% moneyness (at-the-money) with a remaining time to maturity of one year. This procedure is motivated by the work of Cao et al. (2011) and Wagner (2008) who find that the using option implied volatilies improves the fit of the CG model estimated spreads to market spreads in comparison to the CG model that uses an historical volatility estimator. Previous researchers addressed option implied information in linear settings by regressing option market variables on corporate bond credit spreads. They implicitly assume that the relation of equity volatility is inherently linear to observable market spreads (see e.g. Cremers et al. (2008) or Cao, Yu, and Zhong (2010)). My study design deviates from these studies by accounting for non-linear dependencies of forward-looking option market information and resulting CDS spreads within the CG model framework. My work also relates to the study of Stamicar and Finger (2006) who consider a case-study research design and price CDS written on Vivendi, General Motors, Ford and Boeing. However, my study incorporates a larger sample of reference entities and hence provides a higher degree of cross-sectional granularity using option implied volatilities. Stamicar and Finger (2006) conclude that using option implied volatilities improves the performance of the model, especially in crisis times, by providing credit warning signals backed out of option markets.

The parameters $\bar{L}$, R and λ cannot directly be observed in the market and therefore lead to difficulties in the model calibration process. Finger et al. (2002) assess the estimation of the global average recovery rate $\bar{L}$ by relying on the proprietary Portfolio Management Data and Standard & Poor's database also considered by Hu and Lawrence (2000). This database contains historical recovery rates for 300 non-financial US firms issued bonds and loans that defaulted in the period 1987 to 1997. From this data, Hu and Lawrence (2000) estimate $\bar{L} = 0.5$ and $\lambda = 0.3$. Since these estimates seem realistic and are by definition time invariant and global, I proceed by applying the same estimates in my study considering European non-financial firms.

The calibration of R poses the difficulty, that the recovery rate on the specific underlying bond cannot be observed, since default of the bond did not (yet) occur. The authors of the original model and researchers like Yu (2006) use the methodology of leaving R as a free parameter and calibrate the model to market spreads by minimizing the pricing errors of some pre-defined goodness-of-fit measure G. For this purpose I choose the squared pricing deviations as goodness-of-fit measure defined as:

$$\min_R G = \sum_{i=1}^{n}\left(s_i^{Model} - s_i^{Empirical}\right)^2. \tag{37}$$

Using this methodology and minimizing G over the first 25 trading days, I find unreasonable values for R most of the time. As a consequence I follow Yu (2006) who finds similar values and thus sets $R = 0.5$ leaving $\bar{L}$ as the free parameter to fit the model. However, following Yu (2006), I also find unreasonable values for $\bar{L}$. Consequently, I fit the model to sector specific recovery rates as given in Gambetti, Gauthier, and Vrins (2019). They publish sector recovery rates extracted from the proprietary Moody's Analytics Default and Recovery Database. I provide more recovery rate specific details in chapter 5.

The risk-free rate was initially considered to be the short end of the Euro yield curve using the 3-month EURIBOR offered rate in line with recommendations in the existing literature for selecting a Euro risk-free reference asset (Remolona, Wooldridge, et al., 2003). When I have employed this measure as a risk-free rate, the model sometimes did not produce any values. Upon closer examination, equation (28)

demands that the risk-free rate embedded into $z = \sqrt{\frac{1}{4} + \frac{2r}{\sigma^2}}$ does not necessarily have to be positive, but cannot be arbitrarily negative, such that the expression in the square root becomes negative. As this is the case for some observations, I have to choose an alternative risk-free rate that exhibits an overall higher interest rate level. A reasonable alternative is to select the US government bond rate since the general interest rate level is typically higher for the US market than for the European market mainly due to the European sovereign debt crisis and resulting monetary policy measures taken by the ECB. Since in the academic literature the US rate is typically used as a reference risk-free asset, I opt for using the 5 year constant maturity treasury bond middle rate to match the duration on constant 5 year maturity CDS contracts as described in chapter 5 in more detail. My selection of the risk-free asset poses the contextual difficulty that the geographical location does not match. On the other hand, it can be argued that international capital mobility can be taken as a given and thus European market participants can obtain exposure to the US risk-free asset, providing economical justification of employing the US government rate as the reference risk-free rate.

4.3. Correlation and Pricing Analysis

After calibrating the model to market observables using historical equity and option implied volatility as a proxy for asset volatility, I assess the pricing performance of each model specification by means of a correlation analysis and frequently cited forecast metrics. Similar to the study of Bedendo et al. (2011), the correlation analysis of model and empirical spreads is conditioned on sectors in order to obtain a sector specific view on the pricing performance of each model specification. I choose parametric as well as non-parametric methods for measuring the relationship between market and model spreads to capture linear as well as non-linear relationships. First, I calculate the Pearson-product moment correlation coefficient for each time series. In order to obtain average sector specific correlations, I transform individual Pearsons r to the Fisher z-transform by $Z = \frac{1}{2} ln\left(\frac{1+r}{1-r}\right)$. This procedure boils down to normalizing the sampling distribution with the aim to mitigate the effect of distribution skew to be able to take the average of all sector specific z-transformed $\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$. Afterwards, the average z-transformed variable is converted back using $\bar{r} = \frac{e^{2\bar{Z}}-1}{e^{2\bar{Z}}+1}$ to an average Pearson product-moment correlation coefficient. This way of obtaining average sector specific correlations appears to posses superior reliability compared to taking the simple time series average as proposed by Hunter and Schmidt (1990) (see e.g. Corey, Dunlap, and Burke (1998)). This transformation step is then performed for the Spearman rank-order correlation to capture possible monotonic non-linear dependencies.

I proceed by assessing possible autocorrelation within the empirical as well as calculated model CDS spreads for each obligor in the spirit of Byström (2006) to investigate whether my sample of reference entities is subject to inefficiencies in the CDS market. My methodology contributes to the study of Byström (2006) by analysing single name CDS instead of considering CDS indices as well as covering different time periods and obligors. Furthermore, I consider autocorrelations in model CDS spreads when using option implied volatility and thus again contribute to the study of Byström (2006). I calculate Ljung-Box test statistics with 5 and 10 lags respectively, closely following the recommendation of choosing $ln(T)$ as the optimal choice of lags to obtain good power properties in finite samples.

For the performance analysis, I calculate three forecasting metrics: Bias, the Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE). The bias for the time series i and number of observations n is defined as the average error following:

$$bias_i = \frac{1}{n}\sum_n e_{i,t} \quad \text{with} \quad e_{i,t} = s_{i,t}^{Model} - s_{i,t}^{Empirical}$$

and captures the general tendency of the model to deliver feasible estimates. However, since negative and positive deviations can offset each other, a model can provide a very low bias but be imprecise at the same time. Thus, I also calculate the mean absolute error that corrects this issue by taking the mean of the absolute values of the errors. It is calculated as follows:

$$MAE_i = \frac{1}{n}\sum_n |e_{i,t}|.$$

In addition, I calculte the root mean squared error (RMSE) as it provides information on the standard deviation of the error and thus is a good model accuracy measure. In my study the RMSE is calculated as follows:

$$RMSE_i = \sqrt{\frac{1}{n}\sum e_{i,t}^2}.$$

In order to obtain sector cross sectional pricing metrics, I average the metric for each time series i within one sector. The performance of each model specification for each sector and time period can then be assessed and compared to other studies that deal with the pricing performance of the CG model. This comparison provides insight on how the model behaves for European CDS spreads accordingly.

4.4. Multi-factor Model CDS Determinants Estimation Procedure

After assessing the model performance, I proceed with analyzing the gap between theoretical and market CDS spreads using a regression-based analysis. The study I use as reference point in this section is published by Bedendo et al. (2011). They analyse the resulting gap by running panel regressions for each sector on weekly deltas in the difference between market and model CDS spreads. Following Bedendo et al. (2011), I calculate weekly gaps for both the standard CG model calibrated with 1000-day equity volatility and the calibrated CG model using option implied volatilities. Regarding the observation frequency, I opt for weekly

observations, since daily observations could be too noisy and would enable possibilities for statistical inconsistencies and distorted estimates. Then I employ a one-way fixed effects panel regression model with the obligor as the reference index. A one-way fixed effects model is used here, since I assume that there exists obligor specific fixed effects that are captured by the individual-specific error component. Thus, I am able to account for unobserved effects that are unique to individuals, but are also time invariant. Furthermore, I do not employ a two-way fixed effects model, because I control for time variant regression estimates by performing panel regressions grouped according my pre-defined time periods covering the financial crisis, the European sovereign debt crisis, the tranquil period and the corona crisis.

I conduct my regression analysis in three different ways. My first approach considers the gap between model and empirical spread as the dependent variable. This approach tries to identify economical drivers of the variation in the error term, hence providing insights into possible miss-specifications and shortcomings of the CG Model in its respective specifications as well as reveal what drivers move market and model spreads apart. With this approach I follow Bedendo et al. (2011) using an updated dataset as well as a different geographical regions. These regions are described in more detail in section 5. The second approach considers the CG Model spread itself as the dependent variable. This econometric setting allows to identify common characteristics and variables that contribute to the total explanatory power of the variation in models spreads. The third approach considers the weekly changes in the empirical CDS spread as dependent variable, thus providing additional insights into variables that explain the variation in weekly market spread independent from my model selection. Thus, I also shed light into possible model misspecifications and shortcomings using this approach. For the selection of the independent variables, I follow the the literature discussed in section 3 and employ firm specific (individual) as well as macroeconomic (common) variables in the panel regression framework.

As firm specific variables, I consider the weekly obligor (company) return, weekly equity volatility, the weekly delta of the credit curve expressed as the CDS slope defined as the mid-spread of a 10-year tenor CDS minus the mid-spread of a one-year tenor CDS of the same reference entity and the illiquidity factor derived by Amihud (2002). The Amihud (2002) is defined as follows:

$$ILLIQ = \frac{1}{N}\sum_{t=1}^{T}\frac{|r_t|}{€\ VO_t}, \tag{38}$$

with $|r_t|$ the weekly absolute return and $€\ VO_t$ the weekly equity turnover euro volume from time $t-1$ to time t. The rational behind the inclusion of equity volatility is the connection between the model calibration procedure in section 4.2 and the interconnectedness and spillover effects between equity and credit markets. The same holds true for equity returns, since equity returns provide important information for equity markets with implications for credit markets and its credit quality. The credit curve is incorporated in my regression analysis, because it captures the variation in obtaining or providing credit protection for different tenors, indicating that the spread is dependent on how long the protection will last. The slope of the credit curve captures market sentiment, because when credit curves are upwards sloping, the likelihood of experiencing a credit event increases due to a greater degree of credit quality uncertainty in the future. If the credit curve is flat, the market does not anticipate that the occurrence of the credit event is more probable in the near future, nor the distant future and vice versa. A downward sloping credit curve indicates that the spreads for the near future are higher than for the distant future due to low anticipated survival rates in earlier periods. The common shape of the credit curve is the upward sloping curve as survival rates mostly decreases over time and the probability of the occurrence of a credit event increases. As I calculate weekly changes in the slope of the credit curve, I capture market sentiment dynamics that can contribute to the overall explanation of the dynamics in the European credit derivatives market. Additionally, Bedendo et al. (2011) as well as Bedendo, Cathcart, El-Jahel, and Liesch (2005) show that the inclusion of the slope of the credit curve can provide meaningful contributions in predicting future changes in credit spreads with short forecasting horizons.

The illiquidity factor derived by Amihud (2002) enters the regression specification because it provides information regarding illiquidity constraints on equity markets with possible spillover and contagion effects on credit markets (Bedendo et al. (2011)). Therefore, I calculate the weekly absolute return of the reference entity $|r_t|$ and the corresponding weekly euro turnover volume $€\ VO_t$ to derive the illiquidity measure in equation (38).

Regarding the common macroeconomic variables, I specify the weekly change in the risk-free rate, defined as the 3-month EURIBOR, the weekly change in the slope of the yield curve, defined as the 12 month EURIBOR minus the one month EURIBOR and the weekly change of the VSTOXX index in my regression analysis. The risk-free rate is included in my regression framework because influential studies like the one of Collin-Dufresne and Goldstein (2001) find that changes in the spot rate are important determinants of the credit market. They argue that the spot rate is theoretically motivated by mechanics of the inherent risk-free asset drift firm value process described in chapter 3.1. An increase in the risk-free rate, increases the asset drift and thus the distance-to-default. In consequence it the likelihood of experiencing a credit event decreases. Thereby, most studies that contribute to this field of research include the risk-free interest rate as a baseline determinant in their regression methodology. The variation in the slope of the yield curve also provides additional information on the inherent default probability because Collin-Dufresne and Goldstein (2001) argue the dynamics in the slope of the yield curve can also influence the interest rate level itself, thus also driving the asset value process. Consequently an inverse relationship between

an increase in the slope of the yield curve and a lowering of credit risk in the overall economy emerges (Collin-Dufresne and Goldstein (2001)).

As an additional aspect, I try to capture the effect of counterparty credit risk in my study. This aspect is predominantly motivated by the work of Arora, Gandhi, and Longstaff (2012) who find that counterparty credit risk is significantly priced in the cross-section of non-financial credit default swap spreads, becoming more statistically significant after Lehman Brothers defaulted. To be able to depict counterparty credit risk in my regression framework, I derive and calculate a counterparty credit risk index (CRI) which closely resembles the CDR counterparty risk index first introduced in 2008 by Credit Derivatives Research. The CDR index consists of the largest 15 equally weighted CDS 5 year mid-quotes for CDS sellers such as Barclays Bank, J.P.Morgan and Goldman Sachs and is depicted in basis points. I replicate this index with an updated dataset suitable for the largest CDS dealers selling European credit derivatives such as Commerzbank, Deutsche Bank and others closer described in appendix F. Besides the weekly change in the CRI, I further take the change in the slope of the CRI into account by calculating the weekly change in the difference between 10-year and 1-year mean cross-sectional CDS spreads for all index constituents. The change in the slope acts in the same way it does for the sample entity CDS slope as an indicator for market sentiment, but now focusing on the credit protection seller side. This allows to capture market sentiment not as an aggregate, but segmented into protection seller and protection buyer side.

In order to address possible autocorrelation, I include one week and two-week lags of the dependent variable. This methodology is in line with common regression model specifications found in the literature such as Bedendo et al. (2011). I do not include a broad stock market index due to collinearity issues arising from individual equity returns that are also included as firm-specific explanatory variables.

Finally, my fixed effects regression model in its three approaches regarding the gap, the CG model spread and the empirical CDS spread is estimated as follows:

$$\begin{aligned}\Delta GAP_{it} &= \alpha + \beta_1 \Delta rf_t + \beta_2 r_{it} + \beta_3 VOL_{it} + \beta_4 \Delta SLOPE_{it}^{CDS} \\ &+ \beta_5 ILLIQ_{it} + \beta_6 \Delta CRI_t + \beta_7 \Delta SLOPE_t^{CRI} + \beta_8 \Delta SLOPE_t^{Yield} \\ &+ \beta_9 \Delta VSTOXX_t + \beta_{10} \Delta GAP_{it-1} + \beta_{11} \Delta GAP_{it-2} + u_{it}\end{aligned} \tag{39}$$

$$\begin{aligned}\Delta CG_{it}^{Model} &= \alpha + \beta_1 \Delta rf_t + \beta_2 r_{it} + \beta_3 VOL_{it} + \beta_4 \Delta SLOPE_{it}^{CDS} \\ &+ \beta_5 ILLIQ_{it} + \beta_6 \Delta CRI_t + \beta_7 \Delta SLOPE_t^{CRI} \\ &+ \beta_8 \Delta SLOPE_t^{YIELD} + \beta_9 \Delta VSTOXX_t \\ &+ \beta_{10} \Delta CG_{it-1}^{Model} + \beta_{11} \Delta CG_{it-2}^{Model} + u_{it}\end{aligned} \tag{40}$$

$$\begin{aligned}\Delta CDS_{it} &= \alpha + \beta_1 \Delta rf_t + \beta_2 r_{it} + \beta_3 VOL_{it} + \beta_4 \Delta SLOPE_{it}^{CDS} \\ &+ \beta_5 ILLIQ_{it} + \beta_6 \Delta CRI_t + \beta_7 \Delta SLOPE_t^{CRI} \\ &+ \beta_8 \Delta SLOPE_t^{YIELD} + \beta_9 \Delta VSTOXX_t \\ &+ \beta_{10} \Delta CDS_{it-1}^{Empirical} + \beta_{11} \Delta CDS_{it-2}^{Empirical} + u_{it},\end{aligned} \tag{41}$$

with $u_{it} = \psi_i + v_{it}$, ψ_i indicating the individual-specific effect in the variation regarding the dependent variable which remains constant across time for each individual. v_{it} depicts the stochastic disturbance term that varies across individuals and time. I estimate each equation in (39), (40) and (41) on the sector level to account for sector specific idiosyncratic aspects and to compare the model specifications within my sample clustered by economical activity. Thus, I am able to identify sector specific drivers of the independent variable within each regression setup.

Afterwards I proceed with analyzing the error term of each regression by performing a principal component analysis (PCA) in fashion of Collin-Dufresne et al. (2001), Cremers et al. (2008) and Bedendo et al. (2011) to be able to identify any common variables that are not included in my regression specification and are potentially driving variations in the error term. For this I follow Bedendo et al. (2011), who define the number components used by the cumulative explained variance. They incorporate principal components until the cumulative explained variance reaches a level of more or equal to 80%. This procedure gives a standardized view on how many components drive the remaining variation in the error term and can thus indicate any misspecifications of my regression analysis. I deviate from Bedendo et al. (2011) by not only assessing the error term of empirical CDS spreads, but also on the error term of gap and CG model regressions in equation (39) and (40). With this extension I am able to examine the regression approach at a broader level.

After the main regression analysis and the subsequent PCA is performed, I proceed with testing the predictive ability of the CG model to forecast market spreads by implementing a sector specific fixed effects model that considers the regression of weekly changes in market spreads on weekly changes in model spreads with one and two lags and weekly deltas of market spreads with lag of one week accordingly defined in equation (42). This approach is related to the study of Bedendo et al. (2011), who also consider the forecasting ability of the CG model via this approach. The economical interpretation is based on the empirical finding "that relevant information is usually revealed and absorbed faster in the equity market rather than in the credit market."[7] Studies like Norden and Weber (2009) and Longstaff, Mithal, and Neis (2005) support this idea by providing evidence on lead-lag relationships between equity and credit markets. Thus, CDS spreads estimated from structural credit risk models should

[7] Bedendo et al. (2011), p. 673

be able to exhibit subsequent predictive behavior in the credit default swap market. Predictive regressions are performed as follows:

$$\Delta CDS_{it}^{Empirical} = \alpha + \beta_1 \Delta CDS_{it-1}^{Model} + \beta_2 \Delta CDS_{it-2}^{Model} + \beta_3 \Delta CDS_{it-1}^{Empirical} + u_{it}. \quad (42)$$

Along the regression estimation process described above, I compute diagnostic statistics to ensure that the aspect of stationarity, multicollinearity and heteroscedasticity are addressed properly and regression estimates are not biased. To quantify stationarity, I calculate the unit root test for panel data of Levin, Lin, and Chu (2002) that considers pooling cross-section time series data for testing the unit root hypothesis. This methodology is still in line with my fixed-effects model as I test each regression variable individually for stationarity using the entire dataset. In a second step, I test for multicollinearity by first estimating a pooled panel regression for each fixed-effects panel regression I estimate. As again only the independent regression variables are assessed, the pooled model does not interfere with the subsequent fixed-effects estimation procedure. I calculate the Variance Inflation Factor (VIF), since multicollinearity between the predictors is typically indicated by the predictors VIF value above some pre-defined threshold. The VIF measures the increase in variance of regression coefficients due to collinearity and provides an index measure of the inflation in standard errors emerging from correlations with other predictor variables. A typical threshold that is often cited in the literature is 3, such that I also use a VIF reference value of 3 to identify whether multicollinearity is existent in my regression specifications. Furthermore, I always correct the standard errors for heteroscedasticity and spatial correlation by estimating heteroscedasticity robust standard errors for panel regressions with cross-sectional spatial dependence. I rely on Driscoll and Kraay (1998), who derive a nonparametric covariance matrix estimator that yields heteroscedasticity and autocorrelation consistent estimates for standard errors which are corrected for robustness to forms of any spatial and temporal dependence. The often cited and popular Newey-West standard errors are also feasible for panel data regressions, but do not correct for possible autocorrelation between individuals, which is especially important for panel data regressions. Therefore I report regression estimates with Driscoll-Kraay standard errors throughout my analysis.

5. Data

Since my study is mainly based on CDS spread, return and leverage data, I rely on the database of Refinitiv Datastream. It is the most comprehensive data source for CDS related data and is widely used for various fields of research with an emphasis on equity and credit markets. I choose to rely on Refinitv Datastream data which goes back until the 14th of December 2007. First, I extract European daily individual single-name corporate CDS mid-spreads of 5 year tenors of contracts that include the modified-modified restructuring clause and start reporting on the 14th of December 2007. Additionally I rely on senior unsecured seniorities to have comparable contracts in my sample. My data selection is motivated by the work of Berndt et al. (2007), who show that restructuring risk has material impact on the premium, amounting on average 6% to 8% of the value of protection against non-restructuring default events. I exclude financial firms, since the financial leverage of these firms are not comparable to non-financial firms, which can be traced back to the inherent characteristics of the different business models. The remaining population is afterwards filtered for data quality purposes, such that contracts in my sample report regularly and I exclude the possibility of distorted estimates due to illiquidity reasons. Contracts that stop reporting throughout my observation period are dropped. In general, my sample period covers daily observations spanning the period from 14th of December 2007 until the 14th of December 2020. My final CDS dataset consists of 75 obligors and 3392 observations per obligor, resulting in 254.400 observations. I define eight economical sectors based on the sector classification from Refinitiv Eikon. These sectors consist of the following: Consumer Non-Cyclical, Consumer Cyclicals, Industrials, Utilities, Technology, Healthcare, Energy and Basicmaterials. Furthermore I define the following time periods: *Full Period* (14.12.2007 - 14.12.2020), *Financial Crisis* (14.12.2007 - 31.12.2009), *European Sovereign Debt Crisis* (01.01.2010 - 31.12.2013), *Tranquil Period* (01.01.2014 - 28.02.2020) and the *Corona Crisis* (01.03.2020 - 14.12.2020).

Data for the input variables of the CG model estimation, calibration and subsequent panel regression analysis are also retrieved from Refinitv Datastream. I obtain daily total return index values, actual closing prices, the risk-free rate, the number of common shares outstanding, total financial debt, the daily euro turnover volume, daily VSTOXX index values and minority interest for each obligor. Regarding the derivation of the counterparty credit risk index (CRI), I retrieve CDS quotes in the same fashion as described above for financial institutions, most prominent known to be major European CDS dealers (see appendix F). The sector specific recovery rates are taken from the paper by Gambetti et al. (2019) displayed in appendix G. Since my second specification of the CG model relies on option implied volatilites, I also obtain Refinitiv Datastream option-implied volatilities from constant one year at-the-money put and call options of the respective obligors. As not for all reference entities traded options exist, I define a subsample for my option implied CG model estimation to circumvent this problem. Furthermore, available Refinitv option data does not start reporting on the 14th of December 2007, such that I augment the first period of the financial crisis and hence shortening the observation period to be consistent with the other defined time periods. The subsample period covers the horizon from 1st of January 2010 until the 14th of December 2020 and consists out of 46 obligors with 2857 daily observations totaling to 131.422 data points.

6. Results

6.1. Reference Entity Sample Composition and Summary Statistics

In this section I display my reference entity sample composition of industry sectors, geographical location and credit rating with the aim to provide an extensive overview of my sample. Table 1 gives an overview on the industry sector composition (Panel A) and the country distribution (Panel B) of my sample of CDS reference entities. Panel C depicts the rating distribution of obligors. Considering panel A, it becomes obvious that my sample is distributed across all eight sectors with pronounced weights in consumer non-cyclicals and cyclicals as well as industrials. Together these sectors make up more than half to the entire sample sector distribution. On the other hand the remaining sector weights are rather evenly distributed. Panel B shows that the majority of my sample contracts is either located in Germany, Great Britain or France by marking more than 60% of my sample country distribution. This result is largely driven by the selection process described in section 5, because the most liquid and thus frequently reporting contracts have a tendency for higher data quality and are more likely to pass the imposed data quality filters with a self-selection process tilting the data set towards the European economies. Panel C reveals that the rating distribution is rather homogeneous, because the majority of my sample reference entities belong to the investment grade rating, whereas only roughly 10% can be classified as speculative grade. This fact motivates my study design to not group the CG model estimation and subsequent panel regressions by credit ratings as done in other studies like Collin-Dufresne et al. (2001), but by economical activity and thus sectors. With this classification I am able to identify sector specific idiosyncratic model performance and empirical determinants of credit default swap spreads dynamics.

Since the CG model estimation relies strongly on its input variables, I proceed with providing summary statistics that should allow for identification of sector specific patterns which might be helpful to understand and interpret the model performance and pricing analysis in section 6.2. Furthermore, it allows for an Explorative Data Analysis (EDA) such that the relation between model input variables and the model output can be assessed. Table 2 reports summary statistics for all CG model input variables on a per sector basis. The column named "*mean*" in table 2 depicts the cross-sectional means of the model input variables over the model estimation period. When the columns *mean* and *median* are compared, it becomes obvious that nearly all model input variables exhibit right skewed distributions, except for the energy sector with cross-sectional debt-per-share and stock prices being left skewed. The number of individual variable observations naturally is constant for all sectors with 3392 observations per obligor as my empirical analysis is based on a balanced panel dataset. Only for the variable *option implied volatility*, the number of observations deviates because of the shortened observation window for the option implied volatility CG model estimation explained in chapter 5. Regarding the financial leverage of firms, the sector consumer cyclicals shows the highest leverage with DPS values ranging from 8.17 to 47.15, whereas the sector energy together with technology, utilities and healthcare show low intrasector leverage. When considering equity volatility, it is noticeable that consumer cyclicals exhibit the highest degree of volatility with a mean volatility of 35.42%, whereas in comparison the sector healthcare exhibits on average close to 10 percentage points lower equity volatility. All other sectors show medium volatility with value above 25% but lower than 30%. The comparison between option-implied and equity volatility draws a clear picture. Option-implied volatility shows lower lows and higher highs with overall both lower median and mean values in comparison to equity volatility for all sectors. This indicates that option-implied volatility exhibits a higher degree of sensitivity embedded into derivative markets rather than volatility in equity markets, by also providing a lower overall standard deviation for implied volatility for all sectors. Additionally, the kurtosis of input variables does not show a clear pattern. The values mostly remain smaller than one and alternate in signs between sectors.

This provides evidence on the fact that the data is mainly mesokurtic, except for option-implied volatility which is leptokurtic with kurtosis values above three ranging from 3.2 to 187.7, except for the sectors technology and basic materials. Taking kurtosis, skewness and long right tails into account, the cross-sectional distribution of option-implied volatility is predominantly described by a small number of positive outliers that drive the majority of their time-variation. This supports the idea of quicker absorption of information in derivative markets than in equity markets.

6.2. Correlation and Pricing Analysis

In this section I analyze the CG model performance in its two specifications. In a first step, I estimate each model and visualize its mean cross-sectional performance over all sectors. Figure 2 depicts the estimated mean cross-sectional CDS spread for all sectors using the CG model with an asset volatility estimator of 1000-day rolling window equity volatility. The coordinate axis shows spread in bps, whereas the abscissa axis depicts the time covering my sample period from December 2007 to December 2020. It can be seen that the the model spreads tend to follow the empirical spreads. The CG model is able to produce sensible spreads in the cross-section in size and magnitude. After the peak in 2009 with close to 300 bps resulting from increased credit risk due to the global financial crisis, model spreads tend to stay higher than market spreads. This overprediction is mainly due to the selected moving average window which produces so called *ghost effects* when the window is large enough such that initial shocks remain in the estimated time-series even long after the shock has died out (see e.g. Penza, Bansal, Bansal, and Bansal (2001)). This effect is especially prominent when the shock enters and leaves a system quick, which is the case for the financial crisis due to government bail-out programs and

Table 1: Overview of Industry Sector, Country and Rating Sample Composition.

This table gives an overview on the industry sector composition (Panel A) and the country distribution (Panel B) of my sample of CDS reference entities. Panel C depicts the rating distribution of obligors. Industry sector classifications and country affiliations are retrieved from Refinitiv Eikon. For the rating composition the credit rating masterscale from Standard and Poor's Global Ratings is used.

Panel A: Sector distribution of reference entities			
Sector	Count	Sector	Count
Consumer Non Cyclicals	13 (17.33%)	Healthcare	7 (9.33%)
Consumer Cyclicals	23 (30.67%)	Energy	3 (4.00%)
Industrials	10 (13.33%)	Basicmaterials	7 (9.33%)
Utilities	7 (9.33%)	Technology	5 (6.67%)
Panel B: Country distribution of reference entities			
Country	Count	Country	Count
Germany	14 (18.67%)	Netherlands	7 (9.33%)
Great Britain	20 (26.67%)	Austria	1 (1.33%)
France	17 (22.67%)	Greece	1 (1.33%)
Sweden	4 (5.33%)	Spain	2 (2.67%)
Finland	3 (4.00%)	Austria	1 (1.33%)
Switzerland	1 (1.33%)	Italy	3 (4.00%)
Portugal	1 (1.33%)	Belgium	1 (1.33%)
Panel C: Rating distribution of reference entities			
Rating	Count	Rating	Count
Investment Grade			
AAA	0 (0%)	A	8 (10.67%)
AA+	0 (0%)	A-	11 (14.67%)
AA	0 (0%)	BBB+	16 (21.33%)
AA-	1 (1.33%)	BBB	22 (29.33%)
A+	2 (2.67%)	BBB-	4 (5.33%)
Speculative Grade			
BB+	5 (6.67%)	CCC	0 (0%)
BB	0 (0%)	CC	0 (0%)
B	1 (1.33%)	C	0 (0%)
B-	1 (1.33%)	D	0 (0%)

further economical support, reducing uncertainty directly absorbed in the credit derivatives market as reflected in decreasing market spreads. This behavior of model spreads is mainly due to a mismatch between credit and equity markets, since equity volatility stays higher for longer. This finding matches the study of Bedendo et al. (2011), who show the same behavior of the CG model for the American market. In comparison to the first peak, the model can, on average, cope better with the peaks due to sovereign debt crisis in 2011 and 2012. It yields mean cross-sectional values of 230-250 bps, where in 2013 the model spreads again closely follow market spreads, declining to 80 bps thereafter in the defined tranquil period of 2015. Since credit risk increased gradually over time, the specification of the model allows for a better performance and estimation of spreads. The peak in the corona crisis is again picked up by the model with subsequent higher spreads after the initial peak in March 2020, analog to the peak in the financial crisis, again mainly due to a ghost effect of the selected window length.

Figure 3 shows the estimated mean cross-sectional CDS spread for all sectors using the CG model with an asset volatility estimator of one year at-the-money option-implied volatility. It is visible that the model is in the cross-section able to better fit to short term changes in the market spread, especially noticeable in the two peaks of the sovereign debt crisis 2011 an 2012 with a reduction in market spread between the two peaks closely followed by the model. Spreads are in the short run much more volatile than compared to model spreads in figure 2, supporting the hypothesis that information is quicker absorbed in derivatives markets and thus reflected in its price. This allows for a better fit of the model compared to market spreads. The corona crisis of March 2020 shows that model severely overpredicts market spreads in times of uncertainty and high implied volatilities, leading to a deviation of model and market spreads. In comparison to figure 2, this overprediction shows the short term sensitivity of the model, which is in case of the extreme uncertainty in March 2020 oversensitive, result in a divergence between

Table 2: CG Model Input Summary Statistics.

This table presents summary statistics for all CG model input variables on a sector basis. The variable DPS depicts the debt-per-share of each firm for the sector cross section, estimated as described in chapter 4.2. $Equity$ represents the sector cross-sectional stock price. $EquityVolatility$ denotes the 1000 trading-day moving average standard deviations scaled by $\sqrt{252}$ to obtain annualized volatilities. $Option-impliedVolatility$ depicts the Black-Scholes extracted volatilites of obligor specific at-the-money call and put options with a constant maturity of one year. $Risk-free-rate$ represents the 5 year constant maturity treasury bond middle rate. I report the time-series means of the daily mean (*Mean*), standard deviation (*SD*), skewness (*Skew*), kurtosis (*Kurt*), minimum (*Min*), fifth percentile (*5%*), 25th percentile (*25%*), median (*Median*), 75th percentile (*75%*), 95th percentile (*95%*), and maximum (*Max*) values of the cross-sectional distribution of each variable. The column labeled *n* indicates the number of individual variable observations within each sector.

	n	*Min*	*5%*	*25%*	*Mean*	*Median*	*75%*	*95%*	*Max*	*SD*	*Skew*	*Kurt*
Sector Consumer Non Cyclical:												
Cross-Sectional Distribution												
DPS	3392	12.5509	12.6867	14.3745	18.6058	16.5850	21.7585	30.9187	32.4455	5.8206	0.8447	0.0381
$Equity$	3392	13.0485	17.3355	24.0894	32.3930	31.7985	39.6753	49.1130	55.6892	9.8735	0.2423	-0.5877
$EquityVolatility$	3392	0.2039	0.2151	0.2389	0.2748	0.2655	0.3218	0.3437	0.3509	0.0442	0.2125	-1.1042
$Option-impliedVolatility$	2857	0.1272	0.1805	0.1999	0.2261	0.2190	0.2465	0.2919	0.4712	0.0362	1.6403	18.8232
$Risk-free-rate$	3392	0.0019	0.0041	0.0124	0.0170	0.0166	0.0223	0.0295	0.0373	0.0073	0.1386	-0.4580
Sector Consumer Cyclicals:												
Cross-Sectional Distribution												
DPS	3392	30.4068	30.7789	36.8354	53.4215	49.9450	66.5672	85.9017	86.3647	18.5970	0.5780	-0.0082
$Equity$	3392	11.6734	16.9983	28.8271	43.6177	40.4305	57.4919	78.2983	119.2181	19.8188	0.2901	0.3456
$EquityVolatility$	3392	0.2345	0.2517	0.2820	0.3542	0.3272	0.4511	0.4937	0.5067	0.0867	0.3507	-1.1581
$Option-impliedVolatility$	2857	0.1448	0.2101	0.2379	0.2790	0.2639	0.3061	0.4043	0.6673	0.0612	1.7125	11.9964
$Risk-free-rate$	3392	0.0019	0.0041	0.0124	0.0170	0.0166	0.0223	0.0295	0.0373	0.0073	0.1386	-0.4580
Sector Industrials:												
Cross-Sectional Distribution												
DPS	3392	9.4076	9.4695	11.1595	13.4879	12.9934	15.2042	19.5853	19.7391	3.0362	0.7547	0.0725
$Equity$	3392	7.1427	9.1722	12.4001	20.0576	18.1908	26.0221	36.9305	43.8587	8.9835	0.3339	-0.4849
$EquityVolatility$	3392	0.2198	0.2372	0.2736	0.3252	0.3101	0.3930	0.4216	0.4422	0.0634	0.2618	-0.9924
$Option-impliedVolatility$	2857	0.1354	0.2071	0.2291	0.2608	0.2483	0.2813	0.3651	0.6444	0.0493	2.1614	14.9150
$Risk-free-rate$	3392	0.0019	0.0041	0.0124	0.0170	0.0166	0.0223	0.0295	0.0373	0.0073	0.1386	-0.4580
Sector Utilities:												
Cross-Sectional Distribution												
DPS	3392	7.0131	7.0563	8.6182	10.4272	9.9958	11.5565	14.9362	23.4762	2.9656	1.4362	4.1502
$Equity$	3392	6.0838	7.2960	9.3095	12.3694	11.7405	14.68715	20.7756	25.1797	3.9910	0.7698	0.4068
$EquityVolatility$	3392	0.1786	0.1911	0.2296	0.2726	0.2705	0.3225	0.3565	0.3707	0.0518	0.0000	-0.6144
$Option-impliedVolatility$	2857	0.1303	0.1743	0.2021	0.2319	0.2260	0.2570	0.3053	0.9261	0.0460	5.9045	187.7552
$Risk-free-rate$	3392	0.0019	0.0041	0.0124	0.0170	0.0166	0.0223	0.0295	0.0373	0.0073	0.1386	-0.4580
Sector Technology:												
Cross-Sectional Distribution												
DPS	3392	5.5528	5.5655	6.3523	7.9656	7.6323	9.1944	12.6402	12.6412	2.1152	0.8544	0.3368
$Equity$	3392	4.2458	5.2343	7.2683	9.1282	9.0310	10.7312	13.2526	18.1690	2.5503	0.5372	0.6853
$EquityVolatility$	3392	0.1938	0.2089	0.2473	0.2885	0.2802	0.3479	0.3742	0.3815	0.0557	0.1007	-0.8781
$Option-impliedVolatility$	2857	0.1495	0.1691	0.2010	0.2339	0.2322	0.2619	0.3111	0.4370	0.0442	0.7318	1.5499
$Risk-free-rate$	3392	0.0019	0.0041	0.0124	0.0170	0.0166	0.0223	0.0295	0.0373	0.0073	0.1386	-0.4580
Sector Healthcare:												
Cross-Sectional Distribution												
DPS	3392	8.1712	8.2207	14.7124	24.1644	22.6190	28.3547	45.5704	47.1587	11.9837	0.5914	-0.4910
$Equity$	3392	19.3480	23.4768	30.9402	46.7214	45.0141	60.9503	76.5182	89.3498	17.5787	0.2853	-0.8863
$EquityVolatility$	3392	0.1948	0.2033	0.2243	0.2627	0.2447	0.3025	0.3820	0.3870	0.0547	0.5686	-0.5580
$Option-impliedVolatility$	2857	0.1797	0.1970	0.2156	0.2405	0.2348	0.2578	0.3080	0.4669	0.0350	1.2786	3.2485
$Risk-free-rate$	3392	0.0019	0.0041	0.0124	0.0170	0.0166	0.0223	0.0295	0.0373	0.0073	0.1386	-0.4580
Sector Energy:												
Cross-Sectional Distribution												
DPS	3392	4.9677	4.9989	6.7676	8.0392	8.1252	9.6188	11.3735	11.5489	1.8575	0.1022	-0.9705
$Equity$	3392	9.9433	15.3620	19.4333	21.4967	21.5000	23.5862	26.9941	31.4800	3.2282	-0.2396	1.4324
$EquityVolatility$	3392	0.1800	0.1927	0.2305	0.2632	0.2522	0.3087	0.3395	0.3440	0.0462	0.2782	-0.9757
$Option-impliedVolatility$	2857	0.1526	0.1695	0.1885	0.2193	0.2100	0.2433	0.2943	0.5135	0.0408	1.3767	3.9044
$Risk-free-rate$	3392	0.0019	0.0041	0.0124	0.0170	0.0166	0.0223	0.0295	0.0373	0.0073	0.1386	-0.4580
Sector Basicmaterials:												
Cross-Sectional Distribution												
DPS	3392	12.2274	12.4535	15.7890	19.7934	19.2557	23.2284	33.5750	34.6251	5.9002	0.4108	-0.0740
$Equity$	3392	20.7291	29.2467	40.7976	52.4727	51.4565	64.3536	77.4551	86.2350	14.7322	0.2641	-0.6138
$EquityVolatility$	3392	0.1932	0.2203	0.2438	0.2872	0.2766	0.3433	0.3789	0.3864	0.0536	0.4270	-1.0931
$Option-impliedVolatility$	2857	0.1854	0.2047	0.2255	0.2542	0.2491	0.2761	0.3235	0.4375	0.0372	0.8139	0.9872
$Risk-free-rate$	3392	0.0019	0.0041	0.0124	0.0170	0.0166	0.0223	0.0295	0.0373	0.0073	0.1386	-0.4580

equity and credit derivative markets.

In a second step, I consider the sector specific conditional correlation between market and model spreads for each specification as a preliminary step in the performance analysis. In

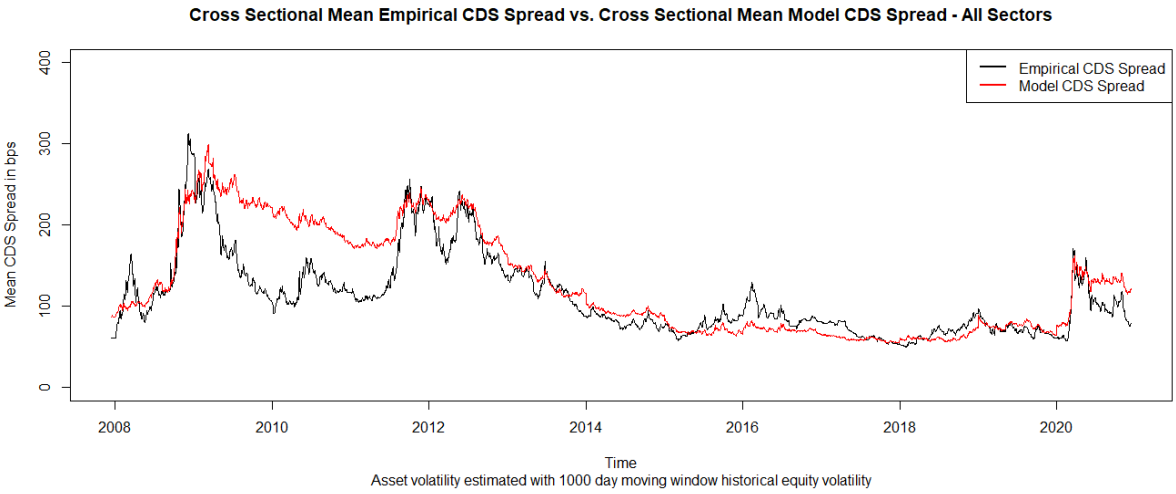

Figure 2: Estimated mean cross-sectional CDS spread for all sectors using the CG model with an asset volatility estimator of 1000-day rolling window equity volatility.

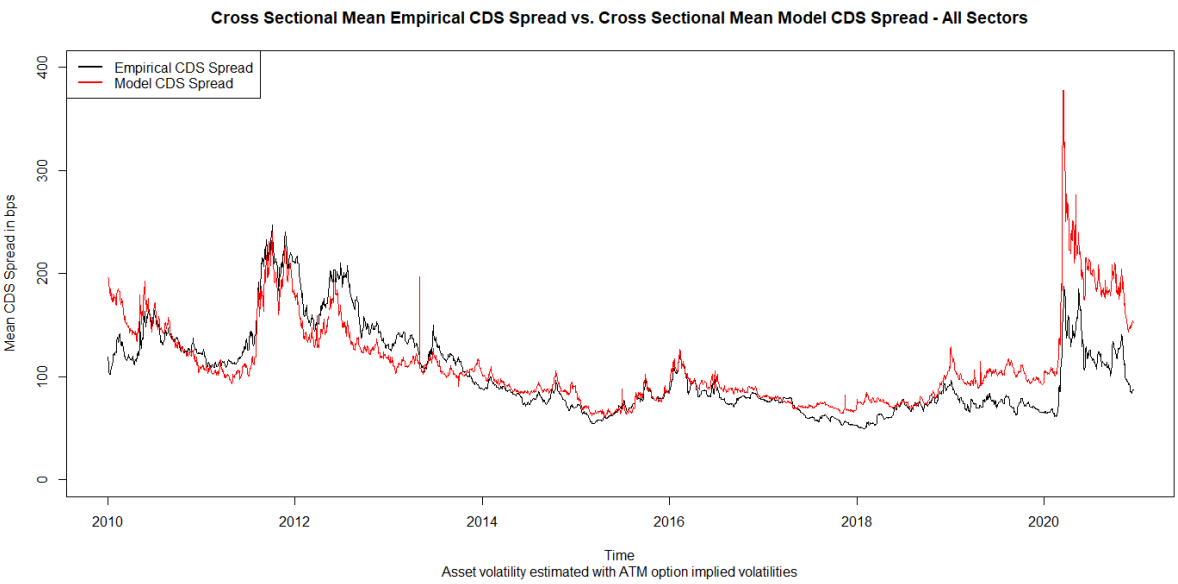

Figure 3: Estimated mean cross-sectional CDS spread for all sectors using the CG model with an asset volatility estimator of one year at-the-money option implied volatility.

table 3 panel A depicts each intrasector correlation estimated with Person product-moment coefficients and panel B depicts the Spearman rank coefficients. For all sectors, except for Energy in its both specifications and technology in its implied volatility specification, the full period exhibits strong correlation between market and model spreads.Values range from 0.546 for the option-implied CG model spreads for the sector consumer non-cyclicals, to 0.814 for the option-implied CG model spreads for the sector utilities.

Regarding the correlation for the historical volatility specifications only, panel A shows that the corona crisis period exhibits the lowest correlations, except for the consumer cyclicals sector, with low to medium correlation of 0.064 for consumer non-cyclical to 0.309 for the sector healthcare. I estimate negative correlations of −0.135 and −0.372 for the sector technology in the financial crisis period and the sector energy for the corona crisis period, indicating that the model is not able to follow market spreads in times of financial turmoil for the respective sectors. If only the implied volatility specification is considered, it is noticeable that model spreads exhibit higher correlation with market spread in times of crisis when volatility is high and spreads shoot up. The tranquil period, on the other hand, shows the lowest correlation for this specification, indicating that the implied vol model is able to follow market spreads best. This can be seen by the significant changes in both credit and equity markets at the same

time. Furthermore, the implied volatility estimated spreads have significantly higher correlation with markets spreads in the corona crisis period than the historical volatility specification for all sectors. This aspect further supports the visual analysis when comparing figures 2 and 3. Regarding the historical volatility specification, the corona crisis exhibits the lowest correlation, both leading to the conclusion that the model is not able to cope with short term increased dynamics in CDS market spreads. Regarding the full sample period, there is no clear pattern which model specification has higher correlation with market spreads. The comparable continuous low correlation for the historical volatility corona crisis period is thereby compensated by changing medium to high correlations in the European sovereign debt crisis period and tranquil period. In correspondence to that, it has to be noted that the full period correlations are not directly comparable, since the financial crisis is not part of the full period for the implied vol specification.

The estimated Spearman rank coefficients in panel B draw the same picture as the Pearson product-moment correlations in panel A with slight different estimates in magnitude. Still they are similar enough to exclude the possibility of biased and distorted results in the subsequent pricing and regression analysis due to non-linearity. All in all, the conditional pricing analysis shows that there exists a strong enough relation between model and market spreads. My results allow for the detailed pricing and regression analysis in order to assess the ability of the CG model to price credit risk in the the cross-section. The previous results are similar in size and magnitude to the correlations measured by Bedendo et al. (2011).

Table 17 in the appendix depicts the assessment of possible autocorrelation in the spirit of Byström (2006). I report Ljung-Box test statistics for each obligor in the sample of empirical and CG model spreads both estimated with historical volatility and option implied volatility. LB(5) indicates the Ljung-Box statistic with 5 lags and LB(10) with 10 lags respectively. For each specification I report the total number of significant variables defined to equal the 5% significance level or higher. Regarding empirical spreads, 54 CDS time series exhibit significant autocorrelation when 5 lags are estimated, resulting in 72% of the sample. When 10 lags are included, 65 CDS time series show significant autocorrelation accounting for 87% of the sample. This result is in line with Byström (2006) who also finds that the vast majority of CDS index market spreads are autocorrelated. Thus I can support his findings also for individual CDS contracts in the European market. Furthermore, I test estimated model spreads and find that the degree of autocorrelation is smaller than for market spreads, however a significant share of the sample shows significant autocorrelation amounting to 63% for 5 lags and 76% for 10 lags in the historical volatility specification and 70% for 5 lags and 78% for 10 lags in the implied volatility specification. This finding deviates from Byström (2006), who concludes that CDS index model spreads do not show high degrees of autocorrelation. Hence, CDS markets could be subject to inefficiencies. Based on these findings, I cannot support this hypothesis as even different model specifications and number of lags do not change the picture dramatically. It is to note that, although the degree of autocorrelation in model spreads is still high, the historical volatility specification yields a lower share of autocorrelation than the market or implied volatility model spreads, indicating that equity markets are less subject to possible inefficiencies if the narrative of Byström (2006) is applied.

Finally, I assess the actual pricing performance of the CG model in its respective specifications by means of forecasting metrics. To do so, I calculate the bias, the mean absolute error and the root mean squared error as defined in chapter 4.3. Table 4 presents the results of the CG model pricing analysis, whereas Panel A depicts key metrics for the 1000-day historical volatility specification and panel B depicts key metrics for the option implied volatility specification. Considering panel A, over all time periods the highest bias, MAE and RMSE are measured in the consumer non-cyclical sector with values of 77.33, 222.09 and 237.80 bps respectively. The model therefore shows the highest degree of mispricing in absolute terms for this sector. Table 15 in the appendix show that this result is not due to high overall spreads which would relativize the mispricing since median sector spreads amount to 74.118 bps. Table 1 depicts low leverage by low debt-per-share median values of 16.586 and comparable low equity volatility of 0.2665, indicating the model shortcomings of reliably estimating spreads if input variables do not exhibit enough variation to drive model inherent credit risk. The consumer cyclicals sector shows the second worst performance with a bias of 77.91 bps, an MAE of 193.3 bps and a RMSE of 235.86 bps. However, when the overall level of market spreads are considered, the picture changes in favor of this sector since it shows the highest median sector spread amounting to 110.999. Thus, a 50% higher market spread level shows good relative model performance compared to the overall market spread level of other sectors. Table 1 again provides insights on how the distribution of model input variables determine model performance, since obligors for the sector consumer cyclicals show the highest median DPS value of 49.9450, median equity value with 40.4305 and highest equity volatility value with 0.3272. This further underlines the initial indication that model performance is highly dependent on input parameter distribution. In contrast to that, the sectors utilities, technology, healthcare, energy and basic materials all exhibit a negative bias showing that the model tends to underestimate market spreads for these sectors. The comparison between the bias and the mean absolute error further indicates a continuous underperformance of the model relative to market spreads, since negative bias values turn by construction to positive MAE values with same magnitude. When again model input parameter distributions are compared in table 1, the drawn picture above remains since DPS, equity and equity volatility exhibit a too small variation to drive model inherent credit risk.

Comparing different time periods, table 4 shows high MAE and RMSE values for the financial crisis and European sovereign debt crisis and corresponding increased variability

Table 3: Conditional Correlation Analysis.

This table reports the conditional sector specific correlation between daily market and model spreads, whereas model spreads are estimated with the CreditGrades model using 1000-trading-day equity volatility (MA Vol) and option implied volatility (Implied Vol). The sector classification is retrieved from Refinitiv Eikon and consists out of Consumer Non-Cyclical, Consumer Cyclicals, Industrials, Utilities, Technology, Healthcare, Energy and Basicmaterials. Furthermore, I report intrasector time dependent correlations for the following time periods: *Full Period* (14.12.2007 - 14.12.2020), *Financial Crisis* (14.12.2007 - 31.12.2009), *European Sovereign Debt Crisis* (01.01.2010 - 31.12.2013), *Tranquil Period* (01.01.2014 - 28.02.2020) and the *Corona Crisis* (01.03.2020 - 14.12.2020). I augment the time period *Financial Crisis* (14.12.2007 - 31.12.2009) for the option-implied volatility CG model specification, due to a smaller subsample and shortened observation window. Panel A presents Pearson product-moment correlations, whereas Panel B reports Spearman rank correlations. To obtain average sector specific correlations, I follow Corey et al. (1998) and transform individual correlations r to the Fisher z-transform via $Z = \frac{1}{2} ln\left(\frac{1+r}{1-r}\right)$. Then the average of all sector specific z-transformed is derived $\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$. Afterwards the average z-transformed variable is then back-converted using $\bar{r} = \frac{e^{2\bar{Z}}-1}{e^{2\bar{Z}}+1}$ to obtain an average sector specific correlation coefficient.

Panel A: Pearson Product-Moment Correlation

	Consumer Non-Cyclical		Consumer Cylclicals		Industrials		Utilities	
	MA Vol	Implied Vol	MA Vol	Implied Vol	MA Vol	Implied Vol	MA Vol	Implied Vol
Full Period	0.596	0.546	0.636	0.586	0.732	0.700	0.702	0.814
Financial Crisis	0.379		0.598		0.383		0.576	
European Sovereign Debt Crisis	0.466	0.516	0.576	0.573	0.555	0.658	0.664	0.831
Tranquil period	0.447	0.488	0.378	0.327	0.583	0.618	0.503	0.675
Corona Crisis	0.064	0.560	0.572	0.709	0.125	0.622	0.308	0.582

	Technology		Healthcare		Energy		Basicmaterials	
	MA Vol	Implied Vol	MA Vol	Implied Vol	MA Vol	Implied Vol	MA Vol	Implied Vol
Full Period	0.571	0.317	0.551	0.634	0.319	0.282	0.685	0.736
Financial Crisis	-0.135		0.315		0.216		0.431	
European Sovereign Debt Crisis	0.611	0.558	0.442	0.484	0.589	0.470	0.660	0.709
Tranquil period	0.594	0.449	0.224	0.383	0.518	0.836	0.660	0.689
Corona Crisis	0.124	0.264	0.309	0.644	-0.372	0.716	0.249	0.765

Panel B: Spearman Rank Correlation

	Consumer Non-Cyclical		Consumer Cylclicals		Industrials		Utilities	
	MA Vol	Implied Vol	MA Vol	Implied Vol	MA Vol	Implied Vol	MA Vol	Implied Vol
Full Period	0.613	0.475	0.745	0.698	0.751	0.764	0.708	0.842
Financial Crisis	0.472		0.637		0.406		0.646	
European Sovereign Debt Crisis	0.528	0.540	0.642	0.629	0.580	0.624	0.665	0.817
Tranquil period	0.479	0.579	0.542	0.562	0.602	0.711	0.562	0.631
Corona Crisis	0.048	0.552	0.553	0.755	0.043	0.625	0.224	0.722

	Technology		Healthcare		Energy		Basicmaterials	
	MA Vol	Implied Vol	MA Vol	Implied Vol	MA Vol	Implied Vol	MA Vol	Implied Vol
Full Period	0.611	0.339	0.625	0.721	0.559	0.429	0.768	0.853
Financial Crisis	-0.087		0.373		0.502		0.472	
European Sovereign Debt Crisis	0.624	0.574	0.466	0.584	0.754	0.497	0.602	0.669
Tranquil period	0.632	0.572	0.189	0.433	0.441	0.706	0.707	0.734
Corona Crisis	0.210	0.110	0.281	0.542	-0.501	0.678	0.157	0.756

of intrasector pricing measures in crisis times depicted by the standard deviation of MAE and RMSE. This finding shows that model spreads from backward-looking equity volatility are subject to the inability to adjust to new information quick enough, such that pricing gaps converge over defined time periods. This implies that the consumer non-cyclicals, consumer cyclicals, healthcare and basicmaterials sectors exhibit the highest forecasting metrics in the financial crisis, whereas the sectors energy, technology, utilities and industrials show highest mispricing in the European sovereign debt crisis. Considering time periods with lowest mispricing, the RMSE seems lowest for the sector consumer cyclicals, industrials and energy, as expected for the tranquil period. Counter intuitively for all remaining sectors, the model performs best in the corona crisis compared to all other time periods, indicating a higher degree of adaptation to dynamical changes in market environments than anticipated.

Turning to panel B, I test the hypothesis that the CG

model calibrated using implied volatility overall performs better than the model in panel A. Furthermore, I test the hypothesis whether the model in panel B outperforms in times of increased volatility, i.e. in crisis periods as indicated by Wagner (2008) and Cao et al. (2011). Panel B illustrates that the consumer non-cyclical sector shows the worst model performance with the highest bias, MAE and RMSE for all time periods considered. If the performance between historical and option implied volatility is regarded, the latter performs worse for the full period in the magnitude of 30 to 40 bps between forecasting measures with corresponding increased variability in pricing measures. This result provides a first conflicting result, in contrast to the findings of Wagner (2008) and Cao et al. (2011). Further assessment shows that my initial findings seem to be highly sector dependent as the consumer cyclicals, industrials, technology and healthcare sectors exhibit a better model performance when forward-looking option-implied volatility is used. This type of volatility reduces the pricing metrics between 20 bps for technology to about 50 bps in the cross-section for the consumer cyclicals sector. Besides that, also the variability of pricing metrics seem to be highly reduced. On the other hand, sectors energy and basicmaterials show worse implied-volatility based model performance besides the already mentioned consumer non-cyclicals sector. Although showing slight worse implied volatility model performance, the utilities sector exhibits decreased variability in the Bias, MAE and RMSE.

When the performance clustered by time periods is assessed between panel A and B, it turns out that for the majority of sectors, the European sovereign debt crisis yields the highest pricing error, consistent with findings of panel A. The corona crisis in panel A exhibits much higher pricing errors than in panel B for the consumer cyclicals and technology sectors with a magnitude of up to 50 bps. If table 1 is considered, the distribution of model specific parameters again explain the pricing performance, since for consumer cyclical the median option implied volatility is the highest among all sectors with 0.2639, closely followed by 0.2322 for technologies. Thus, again the narrative of my pricing analysis continuous in which parameter distributions determine the model performance contingent on model calibration procedures, implying that model performance is by definition inherently sector specific due to unique sector specific model input parameter distributions.

6.3. Model Gap Analysis

In this sub-section, I analyze the model pricing error defined as the gap between model and market spreads in correspondence to section 4.3. For that I implement sector specific panel regressions on obligor specific and individual factors as described in chapter 4.4. In appendix E, I report regression diagnostic statistics valid for all regression setups as staionarity and multicollinearity are analyzed for all regression variables used. The results of the unit root test based on Levin et al. (2002) in table 18 show that all regression variables have p-values of 0, rejecting the null of non-stationarity. The only exception is the weekly change in the slope of the yield curve. However, as Bedendo et al. (2011) employ this variable in their regression procedure and all other variables are stationary, I proceed with the analysis of VIF values to a set threshold of 3. The results are presented in table 19. Panel A shows VIF values for regression variables using historical volatility information, ranging from 1.01 to 2.06 at maximum, hence indicating no multicollinearity concerns for these regressions. Panel B depicts the same range for VIF values, also indicating no evidence of multicollinearity.

The corresponding results for the gap estimated between the CG model calibrated with equity volatility can be found in table 5. Table 5 shows that the adjusted R^2 ranges from 0.066 for the consumer non-cyclicals sector to 0.687 for the technology sector, indicating vast sector specific explanatory power of selected independent variables.

The first finding to discuss is that the weekly change in the counterparty risk index (CRI) exhibits highly significant factor loadings for all sectors with negative signs, such that the weekly change in the gap reduces if the weekly change in CRI is positive. This result is in line with intuition since model and market spreads should converge if counterparty credit risk increases, thus driving the aggregated level of credit risk higher and therefore increasing model as well as market spreads. Furthermore, counterparty credit risk seems to be a highly important determinant of the gap, that is not controlled for in the regression scope of other empirical studies. The factor loadings for the weekly change in the VSTOXX index also reflect this aspect. It captures the overall level of risk and risk aversion and shows that estimated coefficients are comparable in size and magnitude, even though they are only statistically significant for sector regressions (1), (2), (3), (7) and (8). The weekly change in the slope of the CRI also has comparable significant coefficients in magnitude and size. An increase in the change of the slope will reduce the weekly change in the gap, implying that market sentiment will increase since the difference in long tenor minus short tenor spreads increase. Hence default probabilities in the short run are reduced and default probabilities in the long run are higher due to a technical reduction in long term survival probabilities. The results of the company returns are intuitive and statistically significant at 1% for the majority of sectors, since a negative sign states that an increase in returns has a larger reducing effect on market spreads than on model spreads, supporting the findings of Bedendo et al. (2011). Equity volatility estimates yield mixed results with alternating signs for the sectors and only three sectors show significant coefficient estimates. This results contradicts the study of Bedendo et al. (2011) who control for option implied volatility and find that volatility is a crucial determinant of the gap and CDS market and model spreads. Contributing to this result, I find that changes in the credit curve also yield mixed results with significant estimates for half of all sectors. Hence, the credit curve does seem to contribute to the dynamics of the gap, but it does not correspond with the effect described in chapter 4. This result is in line with Bedendo et al. (2011) who confirm this ambiguity for the North Ameri-

Table 4: Model Pricing Analysis.

This table presents the results of the CG model pricing analysis. Panel A depicts key metrics for the 1000-day historical volatility specification and panel B for the option-implied volatility specification. I divide my sample into 8 sectors and 5 time periods. The sector classification is retrieved from Refinitiv Eikon and consists of: Consumer Non-Cyclical, Consumer Cyclicals, Industrials, Utilities, Technology, Healthcare, Energy and Basicmaterials. I further divde my sample according to the following time periods: *Full Period* (14.12.2007 - 14.12.2020), *Financial Crisis* (14.12.2007 - 31.12.2009), *European Sovereign Debt Crisis* (01.01.2010 - 31.12.2013), *Tranquil Period* (01.01.2014 - 28.02.2020) and the *Corona Crisis* (01.03.2020 - 14.12.2020). Since my option implied CG model specifications consists out of a smaller sub-sample with shortened observation window, I augment the time period of the financial crisis to ensure comparability within each time period as described in chapter 5. I report the following three pricing metrics and their corresponding standard deviations: the bias (*Bias*), the mean absolute error (*MAE*) and the root mean squared error (*RMSE*). Their estimation procedure is described in chapter 4.3.

	Bias	*Std(Bias)*	*MAE*	*Std(MAE)*	*RMSE*	*Std(RMSE)*
Panel A: Historical Volatility Specification						
Sector Consumer Non Cyclical						
Full Period	77.33	156.10	222.09	144.06	237.50	145.90
Financial Crisis	107.79	211.70	297.97	195.93	309.47	195.87
European Sovereign Debt Crisis	74.73	151.34	215.89	139.59	218.87	140.18
Tranquil Period	72.65	145.38	205.68	134.35	212.84	134.89
Corona Crisis	47.92	126.21	184.57	115.26	188.58	115.80
Sector Consumer Cyclicals						
Full Period	77.91	53.34	193.30	42.93	235.86	49.92
Financial Crisis	87.28	92.49	291.31	73.86	327.87	77.34
European Sovereign Debt Crisis	174.24	71.50	255.84	60.25	272.86	62.78
Tranquil Period	12.62	38.49	119.48	29.69	126.90	30.65
Corona Crisis	75.60	60.07	198.40	46.29	210.39	48.05
Sector Industrials						
Full Period	6.59	35.18	86.91	25.71	100.77	27.43
Financial Crisis	17.35	49.26	112.99	36.24	130.50	35.38
European Sovereign Debt Crisis	23.35	47.49	105.01	35.77	114.16	36.44
Tranquil Period	-8.09	25.48	67.82	16.34	74.65	17.45
Corona Crisis	8.44	35.13	76.52	26.68	81.61	26.82
Sector Utilities						
Full Period	-86.60	17.90	86.99	17.73	109.30	26.25
Financial Crisis	-76.70	13.77	76.70	13.77	92.90	20.17
European Sovereign Debt Crisis	-143.57	39.99	143.59	39.99	160.19	47.13
Tranquil Period	-59.61	9.86	59.70	9.8	63.50	10.48
Corona Crisis	-34.60	10.93	40.17	7.07	41.26	7.18
Sector Technology						
Full Period	-61.68	34.25	77.33	33.48	129.55	77.21
Financial Crisis	-51.43	14.29	65.31	5.93	75.14	6.85
European Sovereign Debt Crisis	-115.19	95.76	140.67	89.52	200.02	143.96
Tranquil Period	-38.19	10.47	44.59	13.58	50.01	15.96
Corona Crisis	-0.73	28.18	43.41	18.19	49.25	20.16
Sector Healthcare						
Full Period	-14.61	21.18	69.13	22.56	97.00	40.72
Financial Crisis	74.33	115.97	170.39	98.06	183.91	99.87
European Sovereign Debt Crisis	-23.74	19.83	75.49	17.70	88.41	26.71
Tranquil Period	-37.33	5.87	37.33	5.87	39.03	6.03
Corona Crisis	-22.04	2.45	22.26	2.36	24.55	3.03
Sector Energy						
Full Period	-63.53	8.59	63.55	8.59	73.63	11.40
Financial Crisis	-51.74	2.61	51.74	2.61	56.62	3.38
European Sovereign Debt Crisis	-85.23	16.32	85.23	16.32	97.51	21.37
Tranquil Period	-55.84	7.85	55.84	7.85	59.44	7.98
Corona Crisis	-44.41	4.87	44.71	5.16	60.73	8.05

(Continued)

Table 4—continued

Sector Basicmaterials						
Full Period	-75.69	15.49	76.44	15.75	88.63	20.32
Financial Crisis	-130.32	42.46	130.32	42.46	139.22	44.34
European Sovereign Debt Crisis	-81.05	19.56	83.43	18.97	90.07	19.02
Tranquil Period	-58.65	9.59	58.70	9.59	62.67	10.89
Corona Crisis	-39.56	4.34	39.56	4.34	43.08	4.35
Panel B: Option-Implied Volatility Specification						
Sector Consumer Non-Cyclical						
Full Period	132.01	207.65	266.27	190.90	274.90	191.47
European Sovereign Debt Crisis	129.18	216.18	283.44	196.92	286.88	197.86
Tranquil Period	136.05	205.31	258.34	190.06	265.01	190.73
Corona Crisis	114.74	185.76	241.21	169.95	245.81	170.57
Sector Consumer Cyclicals						
Full Period	27.69	61.34	158.70	44.03	184.50	46.67
European Sovereign Debt Crisis	10.20	69.40	179.47	48.64	193.28	50.02
Tranquil Period	22.92	56.15	134.69	42.41	144.59	44.10
Corona Crisis	153.57	90.42	240.96	77.35	264.73	78.61
Sector Industrials						
Full Period	-19.35	38.99	86.75	24.14	96.86	27.14
European Sovereign Debt Crisis	-24.16	52.84	110.79	33.69	117.36	36.00
Tranquil Period	-25.81	30.07	70.71	16.92	74.78	18.00
Corona Crisis	55.49	45.38	90.17	40.47	106.98	43.36
Sector Utilities						
Full Period	-107.26	7.04	107.67	7.33	129.05	9.16
European Sovereign Debt Crisis	-177.08	12.11	177.08	12.11	190.33	12.62
Tranquil Period	-70.12	4.44	70.12	4.44	72.19	4.18
Corona Crisis	-43.65	3.53	49.34	9.19	72.60	31.98
Sector Technology						
Full Period	-30.74	13.00	54.26	11.83	62.85	11.02
European Sovereign Debt Crisis	-50.19	15.81	65.84	7.23	70.50	5.87
Tranquil Period	-33.45	16.97	42.04	13.24	47.15	12.95
Corona Crisis	88.87	30.37	91.00	28.78	104.29	29.90
Sector Healthcare						
Full Period	-31.57	15.69	63.64	21.74	81.34	30.88
European Sovereign Debt Crisis	-23.12	27.92	106.89	40.99	121.47	49.88
Tranquil Period	-39.76	13.64	40.75	12.76	43.29	12.56
Corona Crisis	-10.45	4.85	23.36	2.70	33.46	7.37
Sector Energy						
Full Period	-63.38	18.28	64.49	18.44	71.29	21.18
European Sovereign Debt Crisis	-87.99	28.68	87.99	28.68	93.24	31.32
Tranquil Period	-53.32	13.69	53.32	13.69	55.84	13.80
Corona Crisis	-17.29	1.54	32.69	3.69	43.47	3.24
Sector Basicmaterials						
Full Period	-97.5	26.02	97.52	26.03	109.59	31.27
European Sovereign Debt Crisis	-148.51	46.29	148.51	46.29	151.84	46.93
Tranquil Period	-71.37	16.45	71.37	16.45	77.17	18.92
Corona Crisis	-43.21	6.96	43.53	7.12	45.10	7.38

can credit derivatives market. As the risk-free rate is also a model input parameter, the interpretation of its contribution in explaining the changes in the gap is not straight forward. As in Bedendo et al. (2011), it does not seem to be a determinant of the gap dynamics. The same result holds for the slope of the yield curve. This slope is statistically significant at the 10% level in model (1) and thus provides no contribution to explaining the dynamics of the gap. Illiquidity does not seem to have any significant factor loadings and thus also does not contribute to the explanatory power of the gap. This result is

also in line with Bedendo et al. (2011) who discover mixed results and mostly insignificant regression estimates. However the finding contradicts studies like the one of Tang and Yan (2007) who find significant liquidity risk premia in credit derivatives markets. I conclude that for European CDS gaps, liquidity spillover from equity markets do not seem to affect the difference in market and model spreads. As Bedendo et al. (2011) point out, illiquidity spillover might affect market and model spreads at the same time and hence does not affect the gap. In order to test if illiquidity affects market or model spreads individually, I run subsequent panel regressions on model and market spreads separately in subsection 6.4 and 6.5. The results related to lagged changes in the gap are inconclusive. Most of the coefficients are not significant and vary in sign. Thus, I cannot say some type of mean reverting behavior exists and therefore large changes in the gap are reabsorbed in the following gap spreads. In consequence "convergence is more likely to occur when model spreads move back towards market spreads"[8]. As I estimate regression coefficients with Driscoll-Kraay standard errors correcting for spatial and intertemporal dependence, it is to note that this could also affect the significance of lagged regressions variables.

To analyse the panel regression specification and the completeness of the model, table 5 depicts the number of principal components extracted from the error term that explain 80% or more of the explained variation. It becomes apparent that common factors are largely sector specific since the number of extracted variables is heterogeneous. Common factors explaining 80% of the variation range from 2 in case of regressions (4) and (6), to 9 factors for regression (3). In case of regression (2), (7) and (8) 4 factors are extracted and regression (5) finds 5 principal components that explain at least 80% in the left over variation of the error term. The principal component analysis for the gap error term points out that the regression specification is highly sector specific with sector common variables missing for model (4) and (6), when defining up to two components as model misspecifications. My results are comparable to those of Collin-Dufresne et al. (2001) who also concludes that common factors are still missing from regression specifications. Nevertheless, they do not analyse the gap between some model and market spreads, but address principal components extracted from regressions with the market spread as dependent variable.

Table 6 shows regression estimates for the same setting as table 5, where now the gap is estimated using option-implied CG model spreads. All in all, table 6 allows for similar conclusions as table 5. Counterparty credit risk and equity returns are still a highly significant determinant of the gap, whereas the CRI slope has much weaker explanatory power in the cross-section of sectors. Since forward-looking information backed out of option markets are already embedded in model spreads, further variables controlling for market sentiment might not exhibit the same importance as compared to using historical information for derivation of the model spreads in table 5. The same holds for the slope of the credit curve, which is also less significant in the cross-section as compared to table 5. When considering the model completeness via PCA, I discover that the model error terms are driven by fewer principal components when imposing the restriction to account for at least 80% of its explained variance. This suggests that the model lacks more common factors than the model specifications in table 5. A feasible explanation could be the lack of variables from option markets such as volatility skew or the weekly change in implied volatility. This aspect is also represented in comparably low adjusted R^2 values ranging from 0.045 to 0.333.

6.4. Model Spread Analysis

To better understand the determinants of the European credit default swap spread dynamics, I run sector panel regressions similar to those of the previous section. Table 7 reports the estimates for the equity volatility calibrated model spreads. The difference between the gap regressions of table 5 and the model regressions of table 7 is that now independent variables are also directly used as input variables to derive the dependent variable.

The estimated regression coefficients for company returns are again highly statistical significant and posses the expected sign. In consequence, the model spreads react inverse to company returns. This is not surprising since the model is build on the assumption that credit risk emerges from the relation between the amount of equity and liabilities (financial leverage). Hence, positive equity returns reduce the risk of experiencing a credit event by construction, since leverage would decrease and ceteris paribus the price for credit protection should fall. The same holds true for equity volatility. Since the model is calibrated using 1000-day equity volatility, volatility increases when the risk of hitting the default barrier increases and thus also the model spreads should increase. This is largely the case for model (1), (3), (4) and (5) as they posses positive estimates that are statistically significant at the 5% level. Another model input parameter is the risk-free rate. Because the asset drift of the company is modeled as described in chapter 4.2, I expect that the asset drift increases when the risk-free rate increases and hence the asset value deviates from the default barrier if all other parameters are held equal. In consequence, negative estimates are expected, because credit risk would subsequently decrease as the distance to default increases. This is also observed in table 7, as I obtain negative coefficients for the majority of sectors. Nevertheless these estimates are not always statistically significant. Only models (5), (7) and (8) exhibit estimates that are also statistically significance at the 5% and 1% level respectively. The weekly change in the slope of the yield curve is only significant for the sector healthcare, indicating that it does not explain the variation in cross-sectional model spreads as claimed by Bedendo et al. (2011). Similar to table 5, the illiquidity factor is not significant except for model (1) but with a high coefficient estimates and a large standard error. Therefore I omit this estimate from my study

[8] Bedendo et al. (2011), p. 667.

Table 5: Determinants of the Gap between Market and Model CDS Spreads - Equity Volatility Calibration.

This table reports the estimates from a panel regressions of weekly changes in the difference between market and CG model 5-year CDS mid spreads calibrated with 1000-day equity volatility of equation (39). (1) reports estimates for the sector Basicmaterials, (2) for the sector Consumer Non-Cyclicals, (3) Consumer Cyclicals, (4) Energy, (5) Industrials, (6) Technology, (7) Utilities and (8) Healthcare. Using equation (39), I regress the weekly gap for each obligor (in basis points) on a sector-specific constant, the weekly change in the risk-free rate defined as the 3-month EURIBOR, the weekly change in the slope of the yield curve defined as the 12 month EURIBOR minus the one-month EURIBOR and the weekly change of the VSTOXX index, the weekly change in the CRI and the weekly delta of the CRI slope defined as the mid-spread of a 10-year tenor CRI index value minus the mid-spread of a one-year tenor CRI index value. As firm specific variables, I consider the weekly obligor (company) return, weekly equity volatility, the weekly delta of the credit curve expressed as the CDS slope defined as the mid-spread of a 10-year tenor CDS minus the mid-spread of a one-year tenor CDS of the same reference entity and the illiquidity factor derived by Amihud (2002) estimated by means of equation (38) and using a rolling one week window. I report Driscoll and Kraay (1998) standard errors that are adjusted for heteroskedaticity and spatial correlation in parentheses below the respective estimate. The adj. R^2 is corrected for the fixed-effect and thus includes the full model. It estimates how much of the variation is explained, since idiosyncratic fixed effects contribute to the explanatory power in the variation of the depended variable. The number of principal components is determined such that the explained variance amounts to 80% or more. The sample period covers December 2007 to December 2020.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ Risk free rate	1.359	0.404	−0.277	−0.266	0.393	−0.925	0.984	−0.156
	(1.139)	(1.261)	(1.878)	(0.996)	(2.156)	(1.106)	(1.519)	(0.601)
Δ Slope yield curve	−0.180*	−0.079	−0.127	0.007	−0.116	−0.061	0.014	0.059
	(0.093)	(0.088)	(0.127)	(0.101)	(0.084)	(0.215)	(0.131)	(0.061)
Δ VSTOXX	−0.214***	−0.219**	−0.323*	−0.019	−0.048	−0.158	−0.263**	−0.170*
	(0.075)	(0.097)	(0.193)	(0.170)	(0.072)	(0.200)	(0.121)	(0.094)
Δ CRI	−0.342***	−0.313***	−0.478***	−0.298***	−0.273***	−0.497***	−0.483***	−0.173***
	(0.059)	(0.050)	(0.097)	(0.060)	(0.037)	(0.082)	(0.064)	(0.034)
Δ Slope CRI	−0.223	−0.390***	−0.562***	−0.375***	−0.259***	−0.423**	−0.459***	−0.162***
	(0.165)	(0.108)	(0.165)	(0.079)	(0.100)	(0.184)	(0.131)	(0.055)
Δ Slope CDS	−0.106	0.110	0.446***	0.186	−0.067	0.561***	0.319**	−0.223*
	(0.185)	(0.076)	(0.059)	(0.240)	(0.061)	(0.042)	(0.130)	(0.116)
Company returns	−0.026	−1.119***	−1.377***	0.445**	−0.815***	−0.533*	0.172	−0.776***
	(0.121)	(0.108)	(0.208)	(0.214)	(0.095)	(0.301)	(0.131)	(0.122)
Equity volatility	−0.042	−0.211***	0.268**	−0.117**	0.001	−0.261	−0.055	−0.003
	(0.043)	(0.071)	(0.109)	(0.048)	(0.041)	(0.214)	(0.063)	(0.073)
ILLIQ	1,745.408	0.035	0.113	11.445	−0.081	15.091	−0.265	−2.497
	(10,798.990)	(0.045)	(0.320)	(58.040)	(0.153)	(12.552)	(0.240)	(2.938)
Δ Gap(-1)	−0.052	−0.062**	0.021	0.180**	0.055**	0.089	0.020	0.007
	(0.049)	(0.030)	(0.021)	(0.074)	(0.025)	(0.075)	(0.046)	(0.060)
Δ Gap(-2)	0.042	−0.034	−0.042*	0.001	0.006	−0.118	−0.050	0.001
	(0.052)	(0.028)	(0.022)	(0.063)	(0.029)	(0.082)	(0.055)	(0.053)
Observations	4,056	8,788	15,548	2,028	7,436	3,380	4,732	4,732
Adjusted R^2	0.153	0.066	0.283	0.272	0.096	0.687	0.209	0.074
No. of Principal Components	3	4	9	2	5	2	4	4

Significance Levels: *p<0.1; **p<0.05; ***p<0.01

as it is most likely driven by data issues for this sector. Upon closer inspection it turns out that the basicmaterials sector possesses rather low illiquidity estimates. Hence the regression estimates for the entire sector are likely distorted by extremely large coefficients that compensate the low values of the independent variable. Since the illiquidity measure is calculated uniformly for all obligors as well as sectors and does appear to provide explanatory power in the remaining models, I leave estimated values in table 7 as they are with the note that the regression coefficient should not be subject to interpretation. I conclude that illiquidity does not explain the variation in the weekly change of the model spread. The results related to the VSTOXX variable are counter-intuitive as their estimates are negative, wherefore the model spreads de-

Table 6: Determinants of the Gap between Market and Model CDS Spreads - Option-Implied Volatility Calibration.

This table reports the estimates from a panel regressions of weekly changes in the difference between market and CG model 5-year CDS mid spreads calibrated with option-implied volatility extracted out of at-the-money put and call options with constant maturity of one year in correspondence to equation 39. (1) reports estimates for the sector Basicmaterials, (2) for the sector Consumer Non-Cyclicals, (3) Consumer Cyclicals, (4) Energy, (5) Industrials, (6) Technology, (7) Utilities and (8) Healthcare. Using equation 39, I regress the weekly gap for each obligor (in basis points) on a sector-specific constant, the weekly change in the risk-free rate defined as the 3-month EURIBOR, the weekly change in the slope of the yield curve defined as the 12 month EURIBOR minus the one-month EURIBOR and the weekly change of the VSTOXX index, the weekly change in the CRI and the weekly delta of the CRI slope defined as the mid-spread of a 10-year tenor CRI index value minus the mid-spread of a one-year tenor CRI index value. As firm specific variables, I use the weekly obligor (company) return, weekly equity volatility, the weekly delta of the credit curve expressed as the CDS slope defined as the mid-spread of a 10-year tenor CDS minus the mid-spread of a one-year tenor CDS of the same reference entity and the illiquidity factor derived by Amihud (2002) estimated by means of equation (38) and using a one week window. I report Driscoll and Kraay (1998) standard errors that are adjusted for heteroskedaticity and spatial correlation. The adj. R^2 is corrected for the fixed effect and thus includes the full model. It estimates how much of the variation is explained, since idiosyncratic fixed effects contribute to the explanatory power in the variation of the depended variable. The number of principal components is determined such that the explained variance amounts to 80% or more. The sample period covers January 2010 to December 2020.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ Risk free rate	−0.548	−1.370	0.384	1.984	−1.710	0.422	−3.007	−0.546
	(1.677)	(2.694)	(4.531)	(3.269)	(4.041)	(3.559)	(4.606)	(1.985)
Δ Slope yield curve	0.061	−0.004	0.202	0.117	0.009	−0.091	0.192	0.284**
	(0.142)	(0.116)	(0.152)	(0.121)	(0.131)	(0.198)	(0.233)	(0.115)
Δ VSTOXX	−0.061	0.079	0.906***	0.553*	0.335	0.263	−0.016	0.537***
	(0.176)	(0.181)	(0.316)	(0.332)	(0.267)	(0.185)	(0.167)	(0.145)
Δ CRI	−0.329***	−0.373***	−0.037	−0.507***	0.079	0.079	−0.745***	−0.039
	(0.117)	(0.065)	(0.113)	(0.080)	(0.157)	(0.083)	(0.125)	(0.062)
Δ Slope CRI	−0.210	−0.179*	−0.053	0.138*	−0.164	−0.041	−0.083	−0.228*
	(0.194)	(0.106)	(0.148)	(0.082)	(0.131)	(0.144)	(0.230)	(0.116)
Δ Slope CDS	−0.065	−0.127	0.010	−0.027	0.002	−0.061	−0.034	−0.040*
	(0.081)	(0.110)	(0.096)	(0.082)	(0.062)	(0.121)	(0.079)	(0.022)
Company returns	−0.101	−1.671***	−0.980***	−0.237	−0.331**	−0.141	−0.072	−0.166**
	(0.083)	(0.178)	(0.151)	(0.165)	(0.155)	(0.330)	(0.090)	(0.084)
Equity volatility	0.052**	−0.043	0.164	0.511***	0.137	0.415	0.060*	0.057
	(0.026)	(0.127)	(0.121)	(0.111)	(0.126)	(0.262)	(0.035)	(0.066)
ILLIQ	3,941.809	−1.158	0.087	−136,224.900	−0.039	−0.122	1.093	608.030
	(15,059.270)	(1.059)	(0.338)	(100,019.500)	(0.049)	(0.806)	(0.937)	(6,935.499)
Δ Gap(-1)	−0.071	−0.035	−0.029	−0.194**	−0.172**	−0.247*	−0.082	0.136***
	(0.044)	(0.036)	(0.085)	(0.081)	(0.068)	(0.148)	(0.052)	(0.052)
Δ Gap(-2)	0.026	−0.036	−0.108**	−0.159**	−0.066	−0.149**	−0.040	−0.084
	(0.050)	(0.031)	(0.055)	(0.075)	(0.071)	(0.067)	(0.057)	(0.086)
Observations	2,276	5,121	7,966	1,138	4,552	1,707	1,707	1,707
Adjusted R^2	0.08	0.067	0.045	0.333	0.057	0.096	0.287	0.114
No. of Principal Components	2	3	5	1	3	2	2	2

Significance Levels: *p<0.1; **p<0.05; ***p<0.01

crease when the weekly change in VSTOXX is positive. This would indicate a decoupling of credit derivatives and equity derivatives markets, such that the aggregated level of risk implied in option prices would not translate into the credit risk indicated by the CG model. It could be the case that the calibration method of the model influences this finding. As a consequence I control for this aspect in my panel regressions with CG model option-implied estimated spreads. The corresponding results can be found in table 8. Considering aggregated risk in more detail, I obtain positive factor loadings for the weekly change in the counterparty credit risk index. These findings are in line with prior expectations, indicating that model spreads increase when counterparty credit risk also increases. The weekly change in the slope of the CRI yields mixed results with mostly insignificant estimates and alternating signs. The regression estimates of the slope

from the credit curve however indicates that a steeper slope is related to the model spreads tending to decrease because the aggregated level of risk in short term tenors reduces in comparison to long term tenors. The CG model is thus able to account for the term structure of credit risk. Lagged variables are statistically significant at the 1% level for the consumer non-cyclicals sector. They carry negative coefficients implying that a mean-reverting behavior indeed exists. However, this effect is not found in other sectors such that its existence should be considered with caution. On the other hand, the principal component analysis of model (1) to (8) shows that the pattern from table 5 remains, although the degree of misspecification is slightly lower because for each model, except for model (4), at least one principal component more is needed to attain 80% explained variance in the error term. As I do not account for further granularity in the PCA by e.g. only considering the first two components, the results could arise from statistical noise or randomness. Hence, there are still common sector specific factors missing. This is also found by Collin-Dufresne et al. (2001) who consider local demand and supply shocks on credit derivative markets as a feasible explanation as this story would fit the sector specific variation in components.

Table 8 reports regression coefficients from fixed-effects panel regressions of weekly deltas in model 5-year CDS mid-spreads estimated by the CG model that is calibrated with option-implied volatility that is extracted from at-the-money put and call options with constant maturity of one year based on equation (40). The most important change in coefficient estimates can be seen for the weekly change in VSTOXX, because now increases in the index are positively related to model spreads. These estimates are statistically significant at the 1% level for model (3), (7) and (8). Model (6) exhibits statistical significance at the 5% level, whereas model (4) is only significant at the 10% level. This shows a fundamental aspect in model spread sensitivities contingent on the model calibration procedure with dependencies on what type of markets information are extracted from, subsequently used to estimate the model. My result are thus in line with prior expectations, supporting the hypothesis that determinants of model spreads are highly dependent on model input variables used. Furthermore, counterparty credit risk seems much more important in the sector cross-section using implied volatility as compared to historical volatility. Now model (1), (3), (5), (6), (7) and (8) show factor loadings that are statically significant at the 1% level. This fact indicates that the model calibrated with forward-looking information is much better in picking up counterparty credit risk than the model that relies on past information by using historical volatility. This result hints at a close alignment between credit derivatives and equity markets since the CRI is measured via market spreads in a way such that increased counterparty risk is anticipated by equity derivatives markets before the risk materializes, increasing model spreads. Related to this, the slope of the CRI is also highly important, since the basicmaterials, utilities and healthcare sectors exhibit statistcally significant loadings at the 1% level with negative estimates. Model spreads decrease when short term spreads for counterparties decrease and the slope therefore increases thus providing lower anticipated counterparty risk in the short term. Furthermore, model spreads lagged by one and two weeks accordingly exhibit statistically significant negative coefficients, indicating a higher degree of mean-reverting behavior, which is consistent with the findings of Bedendo et al. (2011). A possible explanation could be that the reactive nature of option-implied volatilties allows the model to faster absorb higher spreads in the cross-section than it is the case with long run volatility windows.

All in all, the comparison between table 7 and table 8 shows that the calibration method has a substantial impact on the identification of determinants of model spreads. Thus, model calibration procedures and following regression interpretations should be assessed accordingly. They should always be tested for different model specifications in order to determine common factors that drive spreads independent from the model calibration.

6.5. Market Spread Analysis

In the third and last model specification I consider market spreads in order to identify drivers of the variation in spreads independent from selected model specifications and calibration methods. To have consistent regression specifications, I stick with the already employed firm-specific as well as common macro-factors. Table 9 shows estimates from the panel regressions of weekly changes in market 5-year CDS mid spreads using equation (41).

As with the two previous regression model specifications (equation (39) and (40), both the weekly change in the counterparty credit risk index as well as obligor equity returns are strongly statistically significant for all sectors at the 1% level and carrying the expected signs. This result indicates that the CRI as well as equity returns are important determinants in the cross-section of European credit default swap spread dynamics. Changes in market spreads, on the other hand, are not sensitive to changes in the risk-free rate. Here higher risk-neutral asset drifts due to an increase in the risk-free rate with corresponding lower default probabilities do not exist in my sample and sample period. This contradicts the results found by Collin-Dufresne et al. (2001), Longstaff and Schwartz (1995) and Duffee (1998) who focus on corporate bond markets. For the term structure of counterparty credit risk, I estimate positive factor loadings that are statistically significant at the 1% and 5% level for all sectors except basicmaterials and healthcare. This result contrasts previously estimated coefficient, that have been usually negative with comparable significance levels. Market spreads tend to increase when the slope of CRI increases, indicating that the market anticipates future higher counterparty credit risk in long tenor contracts (here 10 years) increasing spreads for long tenors which would again increase the slope, by pricing this in to current market spreads. This result is somewhat surprising, since market tenors are always constant at 5 years, such that future increased counterparty

Table 7: Determinants of CG Model CDS Spreads - Equity Volatility Calibration.

This table reports the estimates from the panel regressions of weekly changes in 5-year CDS mid model spreads estimated by the CG model calibrated with 1000-day rolling window equity volatility. The regression specification is performed based on equation (40). (1) reports estimates for the Basicmaterials sector, (2) for the Consumer Non-Cyclicals sector and (3), (4), (5), (6), (7) and (8) for the sectors Consumer Cyclicals, Energy, Industrials, Technology, Utilities and Healthcare respectively. Using equation (40), I regress the weekly change in model spreads for each obligor (in basis points) on a sector specific constant, the weekly change in the risk-free rate defined as the 3-month EURIBOR, the weekly change in the slope of the yield curve defined as the 12 month EURIBOR minus the one-month EURIBOR and the weekly change of the VSTOXX index, the weekly change in the CRI and the weekly delta of the CRI slope defined as the mid-spread of a 10-year tenor CRI index value minus the mid-spread of a one-year tenor CRI index value. As firm specific variables, I consider the weekly obligor (company) return, weekly equity volatility, the weekly delta of the credit curve expressed as the CDS slope defined as the mid-spread of a 10-year tenor CDS minus the mid-spread of a one-year tenor CDS of the same reference entity and the illiquidity factor derived by Amihud (2002) calculated based on equation (38) over a one week window. I report Driscoll and Kraay (1998) standard errors adjusted for heteroskedaticity and spatial correlation. The adj. R^2 is corrected for the fixed effect and thus includes the full model. It estimates how much of the variation is explained, since idiosyncratic fixed effects contribute to the explanatory power in the variation of the depended variable. The number of principal components is determined such that the explained variance amounts to 80% or more. The sample period starts in December 2007 and end is December 2020.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ Risk free rate	−0.278	0.118	−0.968	0.182	−0.776***	−0.265	−0.485***	−1.287**
	(0.232)	(0.381)	(0.784)	(0.128)	(0.274)	(0.323)	(0.155)	(0.512)
Δ Slope yield curve	0.012	0.056	0.063	0.019	−0.009	−0.046	0.020	0.095**
	(0.028)	(0.044)	(0.071)	(0.016)	(0.039)	(0.046)	(0.014)	(0.045)
Δ VSTOXX	−0.171***	−0.031	−0.290***	−0.066*	−0.157**	−0.319***	−0.020	−0.060
	(0.036)	(0.066)	(0.093)	(0.035)	(0.068)	(0.071)	(0.015)	(0.086)
Δ CRI	0.036***	−0.034*	0.124***	0.011	0.036	0.032*	0.011*	−0.004
	(0.010)	(0.018)	(0.031)	(0.008)	(0.022)	(0.016)	(0.006)	(0.023)
Δ Slope CRI	0.002	−0.036	0.061	0.037**	0.003	0.062	−0.008	−0.113***
	(0.026)	(0.029)	(0.093)	(0.017)	(0.036)	(0.053)	(0.013)	(0.038)
Δ Slope CDS	−0.003	−0.024	−0.039***	0.012	−0.040**	−0.023***	−0.003	−0.068
	(0.018)	(0.017)	(0.010)	(0.011)	(0.017)	(0.007)	(0.014)	(0.075)
Company returns	−0.674***	−1.459***	−2.519***	−0.250***	−1.464***	−1.577***	−0.394***	−0.886***
	(0.054)	(0.096)	(0.164)	(0.045)	(0.069)	(0.145)	(0.037)	(0.126)
Equity volatility	0.066**	−0.085*	0.277***	0.039**	0.091**	0.072	0.001	0.057
	(0.031)	(0.051)	(0.093)	(0.017)	(0.045)	(0.062)	(0.021)	(0.069)
ILLIQ	5,875.112**	0.022	0.251	0.286	0.037	4.496	0.110*	−3.219
	(2,776.769)	(0.040)	(0.301)	(5.319)	(0.159)	(7.294)	(0.057)	(3.321)
Δ CG(-1)	−0.005	−0.106***	0.018	0.058	−0.011	0.134**	0.058*	−0.002
	(0.036)	(0.035)	(0.024)	(0.052)	(0.031)	(0.054)	(0.033)	(0.054)
Δ CG(-2)	0.057*	−0.071***	0.003	0.004	0.016	0.00005	0.058***	0.028
	(0.034)	(0.027)	(0.026)	(0.046)	(0.027)	(0.041)	(0.017)	(0.060)
Observations	4,056	8,788	15,548	2,028	7,436	3,380	4,732	4,732
Adjusted R^2	0.434	0.122	0.351	0.465	0.441	0.560	0.377	0.102
No. of Principal Components	4	3	10	2	6	3	5	5

Significance Levels: *p<0.1; **p<0.05; ***p<0.01

credit risk seems to have market wide spillover effects affecting all tenors and are not picked up by estimated model spreads in table 7 and 8. For the slope of the credit curve I obtain estimates in accordance to table 7 and 8, providing cross-validation that not only model spreads are driven by the term structure of credit risk, but also market spreads. Considering equity volatility, factor loadings for basicmaterials and healthcare are positive and statistically significant at the 10% level, whereas for the energy and industrials sector they are significant at the 1% and 5% level respectively. Hence market spreads react to increased equity volatility, supporting the finidngs of Collin-Dufresne et al. (2001) who state

Table 8: Determinants of CG Model CDS Spreads - Option-Implied Volatility Calibration.

This table reports the estimates from the panel regressions of weekly changes in 5-year CDS mid model spreads estimated by the CG model calibrated with option-implied volatility extracted from at-the-money put and call options with constant maturity of one year based on equation (40). The regression specification is performed according to equation (40). (1) reports estimates for the Basicmaterials sector, (2) for the Consumer Non-Cyclicals sector and (3), (4), (5), (6), (7) and (8) for the sectors Consumer Cyclicals, Energy, Industrials, Technology, Utilities and Healthcare respectively. Using equation (40), I regress the weekly change in model spreads for each obligor (in basis points) on a sector specific constant, the weekly change in the risk-free rate defined as the 3-month EURIBOR, the weekly change in the slope of the yield curve defined as the 12 month EURIBOR minus the one-month EURIBOR and the weekly change of the VSTOXX index, the weekly change in the CRI and the weekly delta of the CRI slope defined as the mid-spread of a 10-year tenor CRI index value minus the mid-spread of a one-year tenor CRI index value. As firm specific variables, I consider the weekly obligor (company) return, weekly equity volatility, the weekly delta of the credit curve expressed as the CDS slope defined as the mid-spread of a 10-year tenor CDS minus the mid-spread of a one-year tenor CDS of the same reference entity and the illiquidity factor derived by Amihud (2002) calculated from equation (38) over a one week rollig window. I report Driscoll and Kraay (1998) standard errors adjusted for heteroskedaticity and spatial correlation. The adj. R^2 is corrected for the fixed-effect and thus includes the full model. It estimates how much of the variation is explained, since idiosyncratic fixed effects contribute to the explanatory power in the variation of the depended variable. The number of principal components is determined such that the explained variance amounts to 80% or more. The sample period covers January 2010 to December 2020.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ Risk free rate	0.615	−2.095	7.399*	4.964	1.877	2.033	−1.201	−0.276
	(1.192)	(1.795)	(4.471)	(3.187)	(4.396)	(3.419)	(1.338)	(1.861)
Δ Slope yield curve	0.002	0.059	0.243	0.107	−0.051	−0.141	−0.076	0.253**
	(0.082)	(0.083)	(0.169)	(0.106)	(0.124)	(0.190)	(0.060)	(0.103)
Δ VSTOXX	0.212	0.260	0.956***	0.629*	0.262	0.446**	0.356***	0.723***
	(0.142)	(0.163)	(0.280)	(0.326)	(0.302)	(0.189)	(0.098)	(0.151)
Δ CRI	0.279***	−0.023	0.890***	−0.020	0.536***	0.397***	0.186***	0.286***
	(0.054)	(0.034)	(0.124)	(0.054)	(0.188)	(0.088)	(0.028)	(0.066)
Δ Slope CRI	−0.245***	−0.115	−0.151	0.188*	−0.143	−0.147	−0.122**	−0.391***
	(0.084)	(0.086)	(0.166)	(0.108)	(0.136)	(0.160)	(0.056)	(0.115)
Δ Slope CDS	−0.015	0.028	0.056	0.109*	0.051	0.059	−0.008	−0.040*
	(0.065)	(0.069)	(0.124)	(0.066)	(0.074)	(0.137)	(0.033)	(0.021)
Company returns	−0.216**	−1.892***	−1.334***	−0.367**	−0.580***	−0.132	−0.100**	−0.102
	(0.098)	(0.187)	(0.187)	(0.145)	(0.175)	(0.317)	(0.051)	(0.074)
Equity volatility	0.054*	0.007	0.205	0.661***	0.227	0.426	−0.059***	0.065
	(0.027)	(0.124)	(0.129)	(0.134)	(0.155)	(0.264)	(0.022)	(0.064)
ILLIQ	−8,510.414	−1.041	−0.133	−115,783.700*	−0.085	−0.164	1.297	−2,155.885
	(13,081.930)	(1.065)	(0.267)	(68,406.250)	(0.066)	(0.769)	(0.933)	(6,388.933)
Δ CG(-1)	−0.167***	−0.040	−0.106	−0.119**	−0.160**	−0.234	−0.110	0.085
	(0.056)	(0.034)	(0.067)	(0.053)	(0.064)	(0.147)	(0.080)	(0.066)
Δ CG(-2)	0.035	−0.054**	−0.002	−0.201***	−0.020	−0.134**	0.011	−0.064
	(0.074)	(0.027)	(0.039)	(0.055)	(0.060)	(0.064)	(0.059)	(0.093)
Observations	2,276	5,121	7,966	1,138	4,552	1,707	1,707	1,707
Adjusted R^2	0.211	0.119	0.195	0.468	0.121	0.135	0.345	0.259
No. of Principal Components	2	3	5	1	3	2	2	2

Significance Levels: *p<0.1; **p<0.05; ***p<0.01

that "the contingent-claims approach implies that the debt claim has features similar to a short position in a put option. Since option values increase with volatility, it follows that this model predicts credit spreads should increase with volatility. This prediction is intuitive: Increased volatility increases the probability of default."[9] Illiquidity does only seem to drive spreads for consumer cyclicals for which I estimate a strongly statistically significant factor loading at the 1% level with expected positive sign and plausible magnitude, such that illiquidity equity spillover effects seem to be sector dependent and also for the CG model not detectable. Lagged market spreads show significant factor loadings primarily in one week lagged spreads with positive coefficient, indicating

[9] Collin-Dufresne et al. (2001), p. 2181

that mean-reverting behavior does not seem to exist in market spreads and thus contrasting the findings from option-implied derived CG model spreads. This indicates a disconnection between equity and credit derivatives markets when assessing the inter-dependencies and persistence of spreads. Section 6.6 investigates this relation in more detail. The adjusted R^2 values are overall significantly higher for market regressions than for model regressions, ranging from 0.154 for the sector consumer non-cyclicals to 0.702 for technologies. The PCA reflects this by 3 or more principal components needed to explain at least 80% in error term variance, except for model (4) and (6).

6.6. Predictive Power of CreditGrades

In the last step of my main analysis I test the predictive ability of the CG model to forecast market spreads. As assessed in section 4.4 and pointed out by Bedendo et al. (2011), empirical evidence exists that new information is typically priced in faster in equity markets than in credit markets. To test this hypothesis for the European market I regress weekly deltas in market spreads on weekly deltas in model spreads with one and two lags as well as weekly deltas of market spreads with a lag of one week accordingly. Since the selection of the number of lags is arbitrary, I stick to the study design of Bedendo et al. (2011) to have comparable estimates. Furthermore, I distinguish the predictive ability of the model in its both specifications to provide additional insights on the source of the possible lead-lag relationship between equity and credit derivatives markets. Panel A of table 10 shows the predictive abilities of the CG model, because the variation in empirical market spreads are significantly driven by current model spreads as well as the first weekly lag in change of model spreads. Especially the sector basicmaterials and industrials exhibit significant predictive patterns up to the second lag in model spreads, providing evidence for sector specific predictive abilities of the model. On the other hand, also market lagged variables are significant for both sectors, such that some degree of market inherent inefficiency could influence the predictive ability of the CG model. This aspect underlines the findings in market regressions of section 6.5, such that inefficiencies relate to autocorrelation assessed in section 6.2. When panel A is compared to panel B, it becomes clear that the CG model calibrated with option-implied volatility is also subject to predictive abilities, although now significant for the first lag in changes of the model spread for basicmaterials and energy. My results are comparable to the study of Bedendo et al. (2011) in the way that coefficient estimates are comparable in size and magnitude as well as the overall adjusted R^2. This supports the study of Bedendo et al. (2011) who conclude that significant noise in the estimation procedure or model misspecifications could result in low R^2 values. Additionally, it is interesting to note that when comparing the adjusted R^2 between panel A and panel B, the sector technology, utilities and healthcare show material differences in explanatory power of predictive regressions as the coefficient of determination increases for sector technology and healthcare by more than 10 percentage points when option-implied volatilities are used in the model estimation procedure, whereas utilities decreases by that amount. This aspect provides further evidence on the fact that the predictive ability of the model is both contingent on sectors as well as on the estimation procedure of the model itself, leaving the conclusion by Byström (2006) and Bedendo et al. (2011) who only assess the predictive ability of the model in one specification with a grain of salt.

7. Robustness

7.1. Analysis for CDS Tenors of one and 10 years

In this section I control my previous results for robustness by accounting for the length of which the credit protection will last, namely the tenor of the credit default swap testing results for 1 and 10 year tenors respectively. For this I retrieve one-year and 10-year CDS mid spreads in the same fashion as outlined in section 5 using Refinitiv Datastream. I then proceed with the estimation of determinants of the gap, market and model spreads as in section 6. Appendix H shows panel regression outputs for one year tenors (H.1) and 10-year tenors (H.2). All in all, results are highly similar for CRI, the risk-free rate and the slope of the yield curve since the latter two are still statistically insignificant and do not contribute to the dynamics of gap, model and market spreads in line with my results using 5-year tenors. The factor CRI matches the significance and estimated regression coefficient in size and magnitude, since I calculate 1 and 10 year CRI index values (see appendix F), matching the duration of dependent variable spreads in the robustness analysis. For company returns, I find significantly larger coefficients, especially for 1 year tenors, for gap, model and market spreads, indicating that 1-year credit protection horizons have substantial impact on the likelihood of experiencing a default event by deviating from the default barrier when equity returns are higher. This fact provides further evidence on the fact that model and market spreads are highly sensitive to equity development, especially when the credit protection entails short horizons. In the same fashion I find increased regression coefficients for the term structure of counterparty credit risk and the credit curve, pronounced for 1 year tenors. This underlines that not only obligor equity returns, but also the aggregated credit risk profile on both the credit protection seller and buyer side is highly important when considering the drivers of the cross-sectional cost of credit risk. On the other hand, long tenors of 10 years do not show increased coefficients, showing that only short horizons are sensitive to pivots and changes in the term structure. The 10-year tenor regression outputs in appendix H.2 show that especially in option-implied CG-model and market spread specifications, the factor loadings for the weekly change in VSTOXX are much higher and most often statistically significant as compared to results in section 6. Thus leading to the conclusion that long credit protection tenors are driven by aggregated market wide implied volatility. Consequently, the horizon of

Table 9: Determinants of Market CDS Spreads.

This table reports the estimates from the panel regressions of weekly changes in market 5-year CDS mid spreads using equation (41). (1) reports estimates for the Basicmaterials sector, (2) for the Consumer Non-Cyclicals sector and (3), (4), (5), (6), (7) and (8) for the sectors Consumer Cyclicals, Energy, Industrials, Technology, Utilities and Healthcare respectively. Using equation (39), I regress the weekly change in market spreads for each obligor (in basis points) on a sector specific constant, the weekly change in the risk-free rate defined as the 3-month EURIBOR, the weekly change in the slope of the yield curve defined as the 12 month EURIBOR minus the one-month EURIBOR and the weekly change of the VSTOXX index, the weekly change in the CRI and the weekly delta of the CRI slope defined as the mid-spread of a 10-year tenor CRI index value minus the mid-spread of a one-year tenor CRI index value. As firm specific variables, I consider the weekly obligor (company) return, weekly equity volatility, the weekly delta of the credit curve expressed as the CDS slope defined as the mid-spread of a 10-year tenor CDS minus the mid-spread of a one-year tenor CDS of the same reference entity and the illiquidity factor derived by Amihud (2002) calculated from equation (38) over a one week rolling window. I report Driscoll and Kraay (1998) standard errors adjusted for heteroskedaticity and spatial correlation. The adj. R^2 is corrected for the fixed effect and thus includes the full model. It estimates how much of the variation is explained, since idiosyncratic fixed effects contribute to the explanatory power in the variation of the depended variable. The number of principal components is determined such that the explained variance amounts to 80% or more. The sample period starts in December 2007 and ends in December 2020.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ Risk free rate	−1.654	−0.409	−0.839	0.431	−1.133	0.622	−1.567	−1.071
	(1.117)	(1.117)	(1.747)	(0.978)	(1.724)	(1.113)	(1.480)	(0.654)
Δ Slope yield curve	0.184**	0.106*	0.183	0.011	0.099	0.002	0.008	0.037
	(0.092)	(0.059)	(0.120)	(0.104)	(0.065)	(0.205)	(0.135)	(0.041)
Δ VSTOXX	0.054	0.249**	0.046	−0.036	−0.042	−0.165	0.240*	0.113***
	(0.068)	(0.106)	(0.220)	(0.185)	(0.098)	(0.228)	(0.123)	(0.042)
Δ CRI	0.377***	0.272***	0.602***	0.308***	0.311***	0.531***	0.494***	0.168***
	(0.055)	(0.045)	(0.091)	(0.060)	(0.038)	(0.082)	(0.066)	(0.025)
Δ Slope CRI	0.226	0.333***	0.621***	0.415***	0.271**	0.489**	0.450***	0.049
	(0.177)	(0.100)	(0.215)	(0.086)	(0.107)	(0.220)	(0.139)	(0.056)
Δ Slope CDS	0.106	−0.101	−0.483***	−0.174	0.038	−0.586***	−0.322**	0.151*
	(0.190)	(0.080)	(0.066)	(0.246)	(0.060)	(0.041)	(0.129)	(0.084)
Company returns	−0.648***	−0.355***	−1.149***	−0.690***	−0.630***	−0.988***	−0.566***	−0.108***
	(0.120)	(0.075)	(0.129)	(0.231)	(0.103)	(0.347)	(0.129)	(0.034)
Equity volatility	0.105*	0.071	0.006	0.155***	0.071**	0.356	0.058	0.057*
	(0.055)	(0.054)	(0.036)	(0.051)	(0.033)	(0.251)	(0.061)	(0.030)
ILLIQ	4,116.700	−0.016	0.138***	−13.209	0.087	−10.320	0.380	−0.594
	(10,010.510)	(0.024)	(0.048)	(58.866)	(0.100)	(9.816)	(0.251)	(1.062)
Δ CDS(-1)	−0.026	0.085*	0.060*	0.188***	0.164***	0.086	0.022	0.022
	(0.038)	(0.045)	(0.033)	(0.068)	(0.029)	(0.073)	(0.044)	(0.033)
Δ CDS(-2)	0.080*	0.050	−0.029	−0.010	0.059**	−0.086	−0.032	0.082*
	(0.044)	(0.044)	(0.039)	(0.059)	(0.030)	(0.074)	(0.054)	(0.044)
Observations	4,056	8,788	15,548	2,028	7,436	3,380	4,732	4,732
Adjusted R^2	0.268	0.154	0.494	0.332	0.263	0.702	0.250	0.196
No. of Principal Components	3	6	11	2	6	2	4	4

Significance Levels: *p<0.1; **p<0.05; ***p<0.01

priced in forward-looking information is much longer than typically anticipated by studies that assess 5-year tenors only. On the other hand, VSTOXX factor loadings for 1 year do not significantly deviate from estimated coefficients in section 6, underlining the presumption of pricing in forward-looking information with long time horizons for gap, model and market spreads.

7.2. Time Period Regressions

Besides the assessment of different tenors, I control my results for different time periods. With this approach I test the hypothesis if factor loadings of determinants in gap, model and market spreads are time dependent, reflecting different drivers of spreads contingent on the economical state like crisis and tranquil periods. Since my sample period covers three crisis times (the financial, sovereign debt and

Table 10: Predictive Regressions.

This table reports the estimates from the predictive panel regressions of weekly changes in market spreads on weekly changes in model spreads with one and two lags and weekly deltas of market spreads with a lag of one week accordingly. Panel A reports predictive regressions using equation (42) according to the CG model calibrated with 1000-day rolling window equity volatility. Panel B reports predictive regressions using equation (42) according to the CG model calibrated with option-implied volatility using constant one year at-the-money put and call options. (1) reports estimates for the Basicmaterials sector, (2) for the Consumer Non-Cyclicals sector and (3), (4), (5), (6), (7) and (8) for the sectors Consumer Cyclicals, Energy, Industrials, Technology, Utilities and Healthcare respectively. I report Driscoll and Kraay (1998) standard errors adjusted for heteroskedaticity and spatial correlation. The adj. R^2 is corrected for the fixed effect and thus includes the full model. It estimates how much of the variation is explained, since idiosyncratic fixed effects contribute to the explanatory power in the variation of the depended variable. The sample period covers the time span form December 2007 to December 2020 for panel A and January 2010 to December 2020 for panel B.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Historical Volatility Specification								
Δ model spread	1.064***	0.061***	0.337***	2.543***	0.353***	1.738***	2.292***	0.065**
	(0.118)	(0.014)	(0.050)	(0.363)	(0.045)	(0.462)	(0.494)	(0.027)
Δ model spread (-1)	0.452***	0.020	0.068**	−0.114	0.057**	−0.128	−0.867*	0.022
	(0.130)	(0.013)	(0.027)	(0.356)	(0.025)	(0.163)	(0.456)	(0.028)
Δ model spread (-2)	0.420**	0.014	0.011	−0.376	0.078***	0.309	0.132	0.025
	(0.165)	(0.011)	(0.033)	(0.260)	(0.027)	(0.208)	(0.277)	(0.018)
Δ market spread (-1)	−0.117**	0.087*	0.066	0.205*	0.124***	0.020	0.034	−0.017
	(0.060)	(0.049)	(0.073)	(0.114)	(0.035)	(0.045)	(0.112)	(0.038)
Observations	4,056	8,788	15,548	2,028	7,436	4,732	3,380	4,732
Adjusted R^2	0.198	0.015	0.102	0.211	0.140	0.088	0.157	0.015
Panel B: Option-Implied Volatility Specification								
Δ model spread	0.533***	0.117***	0.305***	0.308***	0.142***	1.183***	0.049**	0.162***
	(0.075)	(0.024)	(0.046)	(0.032)	(0.033)	(0.247)	(0.020)	(0.041)
Δ model spread (-1)	0.149**	0.001	0.033	0.172***	0.016	−0.057	0.010	−0.016
	(0.060)	(0.015)	(0.041)	(0.043)	(0.016)	(0.204)	(0.008)	(0.026)
Δ model spread (-2)	0.016	0.027	0.019	0.003	0.040*	−0.074	0.009	0.026*
	(0.074)	(0.024)	(0.041)	(0.030)	(0.023)	(0.127)	(0.007)	(0.016)
Δ market spread (-1)	−0.012	0.064	0.058	−0.047	0.132***	−0.062	−0.019	0.009
	(0.057)	(0.058)	(0.125)	(0.071)	(0.050)	(0.060)	(0.084)	(0.064)
Observations	2,276	5,121	7,966	1,138	4,552	1,707	1,707	1,707
Adjusted R^2	0.187	0.046	0.122	0.240	0.155	0.201	0.058	0.304

Significance Levels: *p<0.1; **p<0.05; ***p<0.01

corona crisis) and one tranquil period between 2014 to early 2020, I provide various market conditions in which I test the hypothesis that regression estimates are time dependent and measure the degree of explanatory power contingent on market environment and economical conditions. For this I perform panel regressions analog to section 6 grouped by time periods.

Analog to table 5 and 6, table 11 reports estimates from the panel regressions of weekly changes in the difference between market and CG 5-year model CDS mid spreads calibrated using the 1000-day equity volatility for model (1) to (4). In contrast, model (5) to (7) are calibrated with option-implied volatilities. Model (1) covers the financial crisis, model (2) the European sovereign debt crisis, model (3) the tranquil period between 2014 and early 2020 and model (4) the corona crisis. The same holds true for model (5) to (7), although the financial crisis time period is augmented due to limited option data availability. In the first step, only factor loadings between model (1) and (4) are compared. When considering the weekly change in risk-free rate and the

weekly change in the slope of the yield curve, time period results confirm sector results since no regression estimates are statistically significant at the 5% level or higher. The change in VSTOXX index values are strongly statistically significant for the financial crisis and European sovereign debt crisis with negative signs such that gap spreads converge when aggregate option-implied volatility is high. Nevertheless this finding is different for the crisis times as the coronas crisis exhibits no significant loading. The weekly change in CRI is always statistically significant for all time periods with negative signs and hence provides evidence on the fact that counterparty credit risk is a important determinant of CDS spread gaps. The slope of the CRI is only significant for the financial crisis and corona crisis time period such that in times of highly increased short term uncertainty, unique to those time periods, an increase in the slope will reduce gap spreads. This result shows that when the term structure of counterparty credit risk changes, uncertainty and thus also uncertainty in the counterparties financial health will increase long term spreads and thus increase the slope. Hence, an increase in the slope will move model and market spreads upwards and therefore narrowing the gap. Turning to the credit curve, only crisis periods exhibit positive and statistically significant factor loadings implying that an increase in the slope also increases the gap, resulting in time periods where there is a fundamental shift in credit risk term structure for which the CG model derived spreads deviate from market spreads. Company returns are statistically significant for all time periods and carry negative signs, underlining that the gap reduces if returns increase and the asset value of firms deviates from the default barrier. Equity volatility remains with alternating signs, implying it is unlikely driving the gap. Illiquidity is never significant at the 5% level or higher, supporting my findings from section 5.

For the models (5) to (7) that use option-implied CG model calibration to calculate the gaps, I obtain factor loadings for VSTOXX comparable in significance to models (1) to (4), but now with positive signs. This implies that gaps increase when the aggregated implied volatility increases. The risk-free rate is only significant for the tranquil period, supporting the finding in section 6.3 that it does not seem to systematically drive gap spreads. The same holds true for the slope of the yield curve since only model (5) yields a significant factor loading. Company returns, illiquidity and equity volatility are mostly in line with the main results, providing no additional time dependent idiosyncraticies. The lagged gap variable estimates show that time and calibration specific mean-reverting tendencies exist, supporting the result in 6.3 that high option-implied estimated gap spreads seem to be absorbed in following gap spreads which is indicated by negative coefficients of the lagged change in the gap that is statistically significant at the 1% level in model (5) and (6).

Considering table 12 and model (1) to (4), the risk-free rate, the slope of it and the illiquidity measure are not statistically significant in any time period, indicating that model spreads are not driven by these factors. This supports the empirical evidence of my main analysis. Factor loadings are strongly statistically significant at the 1% level for changes in VSTOXX in crisis periods (model (1), (2) and (4)), and at the 5% level for the tranquil period for which the coefficient estimates are reduced by more than half. Consistent with my sector specific regressions, I estimate negative coefficients implying that model spreads decrease when implied volatility increases. Factor loadings for CRI are statistically significant only for the financial crisis and the corona crisis, indicating that model spreads are mainly driven by counterparty credit risk in times of increased aggregated uncertainty with possible spillover and contagion effects in dealer CDS markets. Furthermore, the slope of the CRI index also reflects this finding. The term structure of credit risk is consistent with prior expectations, such that factor loadings are statistically significant when credit risk materializes in crisis times and not significant for the tranquil period. Signs are negative, indicating that model spreads decrease when the term structure becomes steeper, which in line with the chapter 6.4. Company returns are statistically significant for all periods carrying negative signs and thus supporting the evidence that the CG model is able to capture the decrease in PD's when asset values deviate from the default barrier due to increased equity returns. This finding is valid for every economical condition and environment. Corresponding to that, factor loadings for equity volatility show support for the prior expectations of Collin-Dufresne and Goldstein (2001) by displaying positive and statistically significant estimates for the variation in equity volatility in crisis periods only. This fact highlights that the model captures increased default risk when assets are more likely to hit the barrier due to greater variability in equity. The estimates related to the lagged variables are consistent with the gap time period regressions. Interdependencies do not seem to exist when assessing the model estimated with historical volatility. When considering model (5) to (7), the findings change. Analog to table 8, lagged model spreads seem to cause current model spreads to revert, such that mean reverting behavior is present in all time periods affecting one and two week lagged variables. Model calibration also seems to have impact on CRI, since every time period shows statistically significant factor loadings, indicating that counterparty credit risk seems to be important as also depicted in table 13. Regarding VSTOXX factor loadings, the CG model is now able to react in the assumed direction. Hence, model spreads increase when aggregated implied volatility increases.

In the last step of my robustness analysis I consider time period regressions on market spreads in the same way as for gap and model spreads. Since market regressions are independent from model estimation procedures, the results from models (1) to (4) cover all four time periods. For the risk-free rate and the slope of the yield curve coefficients are for no time period significant at the 5% level or higher. Hence, estimated model spreads capture the driver of the variation well. Regarding autocorrelation contingent on market environment, I estimate a positive regression coefficient at 1% for the financial crisis, such that market spreads tend to increase if prior spreads increased. Market spreads have a higher ten-

Table 11: Determinants of the Gap between Market and Model CDS Spreads - Time Period Regressions.

This table reports the estimates from the panel regressions of weekly changes in the difference between market and CG 5-year model CDS mid spreads calibrated with 1000-day equity volatility in models (1) to (4). In models (5) to (7) the model is calibrated using option-implied volatilities extractedfrom at-the-money put and call options with constant maturity of one year. (1) reports estimates for the financial crisis time period using historical volatility calibration, (2) for the European sovereign debt crisis time period using historical volatility calibration, (3) for the tranquil time period using historical volatility calibration, (4) for the Corona crisis time period using historical volatility calibration. Model (5) reports estimates for the European sovereign debt crisis time period using option-implied volatility calibration, (6) for the tranquil time period using option-implied volatility calibration and model (7) for the Corona crisis time period using option-implied volatility calibration. I regress the weekly gap for each obligor (in basis points) on a time period specific constant, the weekly change in the risk-free rate defined as the 3-month EURIBOR, the weekly change in the slope of the yield curve defined as the 12 month EURIBOR minus the one-month EURIBOR and the weekly change of the VSTOXX index, the weekly change in the CRI and the weekly delta of the CRI slope defined as the mid-spread of a 10-year tenor CRI index value minus the mid-spread of a one-year tenor CRI index value. As firm specific variables, I consider the weekly obligor (company) return, weekly equity volatility, the weekly delta of the credit curve expressed as the CDS slope defined as the mid-spread of a 10-year tenor CDS minus the mid-spread of a one-year tenor CDS of the same reference entity and the illiquidity factor derived by Amihud (2002) calculated from equation (38) over a one week horizons. I report Driscoll and Kraay (1998) standard errors adjusted for heteroskedaticity and spatial correlation. The adj. R^2 is corrected for the fixed effect and thus includes the full model. For model (1) to (4) the sample period covers the time span starting in December 2007 up to December 2020, whereas for model (5) to (7) the sample periods covers January 2010 to December 2020. Hence I expand the financial crisis time period for the option-implied regressions.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Δ Risk free rate	0.738	−2.121	−1.423	−1.802	−2.064	−3.602**	−7.306
	(1.698)	(1.294)	(1.496)	(3.330)	(1.554)	(1.692)	(10.910)
Δ Slope yield curve	−0.856*	0.178	0.093	−0.382	0.375**	0.240	0.601
	(0.450)	(0.181)	(0.150)	(0.576)	(0.180)	(0.230)	(1.611)
Δ VSTOXX	−0.698***	−0.339**	−0.111	0.361	0.462***	0.210***	1.161**
	(0.262)	(0.137)	(0.075)	(0.243)	(0.178)	(0.069)	(0.455)
Δ CRI	−0.310***	−0.422***	−0.568***	−0.808***	−0.185***	−0.352***	−0.011
	(0.080)	(0.056)	(0.105)	(0.056)	(0.072)	(0.120)	(0.208)
Δ Slope CRI	−0.445***	0.098	0.080	−1.000***	0.008	0.125	−2.027**
	(0.136)	(0.089)	(0.120)	(0.364)	(0.087)	(0.156)	(0.925)
Δ Slope CDS	0.260***	0.547***	−0.032	0.503***	−0.049	−0.073***	−0.002
	(0.095)	(0.039)	(0.167)	(0.060)	(0.036)	(0.022)	(0.053)
Company returns	−1.104***	−1.366***	−0.599***	−0.944***	−0.633***	−0.419***	−1.069***
	(0.270)	(0.100)	(0.049)	(0.191)	(0.077)	(0.047)	(0.165)
Equity volatility	0.266**	−0.091	−0.120***	0.085	−0.006	−0.013	0.089*
	(0.119)	(0.114)	(0.042)	(0.063)	(0.064)	(0.040)	(0.049)
ILLIQ	0.156	0.012	0.085*	0.004	0.010	−0.018	−0.303*
	(0.410)	(0.041)	(0.049)	(0.171)	(0.030)	(0.038)	(0.155)
Δ Gap(-1)	0.026	0.062	0.001	0.0001	−0.129***	−0.120	−0.026
	(0.026)	(0.053)	(0.027)	(0.045)	(0.030)	(0.084)	(0.103)
Δ Gap(-2)	−0.038	−0.077	−0.018	−0.066*	−0.030	−0.092***	−0.150*
	(0.039)	(0.055)	(0.023)	(0.040)	(0.024)	(0.033)	(0.081)
Observations	7,800	15,525	24,000	2,925	9,476	14,720	1,794
Adjusted R^2	0.115	0.489	0.223	0.660	0.052	0.059	0.146

Significance Levels: *p<0.1; **p<0.05; ***p<0.01

dency to be more persistent in times of high volatility in line with findings of Bedendo et al. (2011), although not entirely consistent with other crisis times as only the financial crisis exhibits significant estimates. Independent regression vari-

Table 12: Determinants of CG Model CDS Spreads - Time Period Regressions.

This table reports the estimates from the panel regressions of weekly changes in 5-year model CDS mid spreads estimated by the CG model calibrated with 1000-day equity volatility in models (1) to (4). Models (5) to (7) are then calibrated with option-implied volatilities extracted from at-the-money put and call options with constant maturity of one year. (1) reports estimates for the financial crisis time period using historical volatility calibration, (2) for the European sovereign debt crisis time period using historical volatility calibration, (3) for the tranquil time period using historical volatility calibration, (4) for the Corona crisis time period using historical volatility calibration. Model (5) reports estimates for the European sovereign debt crisis time period using option-implied volatility calibration, (6) for the tranquil time period using option-implied volatility calibration and model (7) for the Corona crisis time period using option-implied volatility calibration. I regress the change in weekly model spread of each obligor (in basis points) on a time period specific constant, the weekly change in the risk-free rate defined as the 3-month EURIBOR, the weekly change in the slope of the yield curve defined as the 12 month EURIBOR minus the one-month EURIBOR and the weekly change of the VSTOXX index, the weekly change in the CRI and the weekly delta of the CRI slope defined as the mid-spread of a 10-year tenor CRI index value minus the mid-spread of a one-year tenor CRI index value. As firm specific variables, I consider the weekly obligor (company) return, weekly equity volatility, the weekly delta of the credit curve expressed as the CDS slope defined as the mid-spread of a 10-year tenor CDS minus the mid-spread of a one-year tenor CDS of the same reference entity and the illiquidity factor derived by Amihud (2002) calculated using equation (38) over a one week horizons. I report Driscoll and Kraay (1998) standard errors adjusted for heteroskedaticity and spatial correlation. The adj. R^2 is corrected for the fixed effect and thus includes the full model. For model (1) to (4) the sample period spans the time period from December 2007 to December 2020, whereas for model (5) to (7) the sample period covers January 2010 to December 2020. Hence I expand the financial crisis time period for the option-implied regressions.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Δ Risk free rate	0.096	0.149	−0.449	−0.067	0.490	−3.584**	−0.173
	(0.557)	(0.674)	(0.806)	(0.922)	(1.307)	(1.594)	(6.637)
Δ Slope yield curve	−0.117	−0.012	0.099	0.123	0.180	0.347*	−0.736
	(0.158)	(0.070)	(0.082)	(0.212)	(0.115)	(0.183)	(0.986)
Δ VSTOXX	−0.325***	−0.241***	−0.120**	−0.218***	0.604***	0.345***	1.029**
	(0.120)	(0.048)	(0.053)	(0.047)	(0.134)	(0.112)	(0.379)
Δ CRI	0.064**	0.015	0.039	0.077**	0.306***	0.224***	1.545***
	(0.032)	(0.017)	(0.037)	(0.039)	(0.042)	(0.077)	(0.162)
Δ Slope CRI	0.155***	−0.011	0.043	−0.670***	−0.206***	0.138	−1.672**
	(0.058)	(0.031)	(0.052)	(0.173)	(0.066)	(0.119)	(0.666)
Δ Slope CDS	−0.024	−0.017***	0.008	−0.044***	−0.066**	−0.004	−0.079
	(0.022)	(0.005)	(0.012)	(0.011)	(0.026)	(0.012)	(0.049)
Company returns	−2.013***	−2.226***	−0.975***	−2.104***	−0.845***	−0.479***	−0.946***
	(0.213)	(0.074)	(0.039)	(0.087)	(0.085)	(0.051)	(0.149)
Equity volatility	0.286***	0.131**	0.060	0.106**	0.063	0.0002	0.104**
	(0.096)	(0.057)	(0.049)	(0.044)	(0.062)	(0.043)	(0.047)
ILLIQ	0.367	0.039	0.049	−0.226	−0.037	−0.038	−0.663**
	(0.389)	(0.034)	(0.037)	(0.177)	(0.037)	(0.040)	(0.303)
Δ CG(-1)	−0.005	0.012	0.004	−0.083	−0.119***	−0.109	−0.193***
	(0.040)	(0.026)	(0.027)	(0.095)	(0.035)	(0.085)	(0.073)
Δ CG(-2)	0.015	−0.015	−0.025	−0.070	0.014	−0.094***	−0.063
	(0.046)	(0.026)	(0.026)	(0.053)	(0.029)	(0.031)	(0.044)
Observations	7,800	15,525	24,000	2,925	9,476	14,720	1,794
Adjusted R^2	0.302	0.358	0.346	0.593	0.160	0.090	0.360

Significance Levels: *p<0.1; **p<0.05; ***p<0.01

ables like the risk-free rate, the slope of the yield curve, counterparty credit risk, the slope of it and company returns remain consistent with the sector market regressions in terms of statistical significance and the magnitued of the estimates.

Table 13: Determinants of Market CDS Spreads - Time Period Regressions.

This table reports the estimates from the panel regressions of weekly changes in market 5-year CDS mid spreads. (1) reports estimates for the financial crisis time period, (2) for the European sovereign debt crisis time period, (3) for the tranquil time period and (4) for the Corona crisis time period. I regress the change in weekly market spread for each obligor (in basis points) on a time period specific constant, the weekly change in the risk-free rate defined as the 3-month EURIBOR, the weekly change in the slope of the yield curve defined as the 12 month EURIBOR minus the one-month EURIBOR and the weekly change of the VSTOXX index, the weekly change in the CRI and the weekly delta of the CRI slope defined as the mid-spread of a 10-year tenor CRI index value minus the mid-spread of a one-year tenor CRI index value. As firm specific variables, I consider the weekly obligor (company) return, weekly equity volatility, the weekly delta of the credit curve expressed as the CDS slope defined as the mid-spread of a 10-year tenor CDS minus the mid-spread of a one-year tenor CDS of the same reference entity and the illiquidity factor derived by Amihud (2002) calculated with equation (38) over a one week time horizons. I report Driscoll and Kraay (1998) standard errors adjusted for heteroskedaticity and spatial correlation. The adj. R^2 is corrected for the fixed effect and thus includes the full model. The sample period covers December 2007 to December 2020.

	(1)	(2)	(3)	(4)
Δ Risk free rate	−0.709	2.105*	0.848	1.648
	(1.605)	(1.210)	(1.098)	(2.980)
Δ Slope yield curve	0.616*	−0.180	−0.008	0.479
	(0.374)	(0.153)	(0.132)	(0.457)
Δ VSTOXX	0.444**	0.126	0.001	−0.545**
	(0.179)	(0.135)	(0.058)	(0.259)
Δ CRI	0.375***	0.436***	0.603***	0.866***
	(0.064)	(0.061)	(0.118)	(0.055)
Δ Slope CRI	0.601***	−0.103	−0.032	0.514
	(0.144)	(0.095)	(0.115)	(0.368)
Δ Slope CDS	−0.265***	−0.565***	0.038	−0.549***
	(0.087)	(0.039)	(0.175)	(0.066)
Company returns	−0.926***	−0.843***	−0.378***	−1.148***
	(0.138)	(0.093)	(0.052)	(0.177)
Equity volatility	0.003	0.217**	0.176***	0.008
	(0.037)	(0.099)	(0.068)	(0.059)
ILLIQ	0.191***	0.041	−0.037	−0.213*
	(0.041)	(0.041)	(0.024)	(0.117)
Δ CDS(-1)	0.128***	0.075	0.037	0.006
	(0.042)	(0.061)	(0.040)	(0.046)
Δ CDS(-2)	−0.028	−0.054	0.010	−0.026
	(0.064)	(0.061)	(0.038)	(0.052)
Observations	7,800	15,525	24,000	2,925
Adjusted R^2	0.235	0.603	0.221	0.746

Significance Levels: *p<0.1; **p<0.05; ***p<0.01

From that I conclude that these factors drive spreads regardless of the environment in which obligors operate. The slope of the CDS exhibits negative and strongly statistcally significant coefficient estimates in crisis times, whereas it does not in the tranquil period, thus showing that the term structure of credit risk has impact on market spreads when credit risk is more likely to materialize. In contrast to all other previous regressions, I estimate a positive factor loading for equity illiquidity in the financial crisis time period, showing time varying illiquidity spillover effects from equity to credit markets. This result illustrates the importance of distinguishing time periods and provides feasible arguments that drivers of credit risk are not time invariant, but to a certain extent highly time dependent.

8. Conclusion

With this thesis I tried to shed light onto the dynamics of the European credit derivatives market by assessing the performance of the CreditGrades model used to quantify credit risk in the cross-section of my sample of reference entities. The conditional correlations show that strong relationships between estimated model and market spreads exist but they seem to be highly sector dependent. This points out that model performance has be be addressed on a per sector level. The historical volatility calibrated model cannot follow market spreads in times of financial turmoil mainly due to the ghost effect of constant sliding equity volatility window that prevent the model to incorporate new information quickly. On the other hand, the implied-volatility calibration can better adjust to new information, thus supporting the hypothesis of quicker incorporation of information in equity derivatives markets formulated by Cao et al. (2011), Bedendo et al. (2011) and Wagner (2008). Correlations are higher in times of increased volatility both in equity and credit derivatives markets, together with the pricing analysis of forecasting metrics leading to the conclusion that the model parameter distribution has major impact on the model performance. This fact underlines the model related downside that credit risk is driven by the distance between asset value and default barrier. Thus, sectors which do not exhibit high financial leverage and low equity volatility (implied volatility) show severe model under-performance. Practical model implementation always needs to be assessed together with a model back-testing to analyse the relation between parameter distribution and the ability of the model to derive credit risk in the cross-section. When assessing autocorrelations in model and market spreads, I confirm the findings of Byström (2006) that market inefficiencies do exist. This is because market spreads are highly autocorrelated, resulting in effects on pricing and model performance thereafter. However my findings deviate from Byström (2006) in the regard that also model spreads show high degrees of autocorrelation which is later also addressed in performed panel regressions. Turning to the determinants of the gap between model and market spreads, I discover that structural variables explain variations well since the risk-free rate and the slope of it are mostly statistically insignificant. On the other hand, equity returns and counterparty credit risk are almost always useful to explain the weekly variation in gap, model and market spreads. Factors like aggregated implied volatility, the term structure of counterparty credit risk and the credit curve are highly sector specific in the explanatory value of spreads. The illiquidity factor of Amihud (2002) does not seem to drive CDS spreads, thus I can reject the hypothesis that equity illiquidity spillover effects are reflected in gap, model and market spreads. Determinants of model spreads depend, as expected, on model calibration procedures resulting in the ability of the option-implied volatility to pick up market aggregated implied volatility depicted by positive VSTOXX loadings. Mean-reverting behavior of gap and model spreads found by Bedendo et al. (2011) also depends on model calibration procedures, as option-implied derived model spreads exhibit stronger mean-reversion and historical volatility estimated spreads show higher persistence. With my sample of reference entities I can confirm the predictive ability of the CG model also in the European market as also discovered by Byström (2006) and Bedendo et al. (2011). However, upon closer inspection the predictive ability depends on sector and model calibration procedures such that capital structure arbitrage strategies, exploiting this market inefficiency, have to be chosen carefully. My robustness analysis shows that determinants of gap, model and market spreads are related to CDS tenors and time periods. Short term CDS tenors of one year exhibit higher sensitivities to equity returns, the term structure of counterparty credit risk and the credit curve. For long tenors, VSTOXX appears more important, especially for option-implied gap and model spreads, showing that the horizon of priced in forward-looking information is much longer than anticipated when using 5-year tenors. Panel regression grouped by time periods show that determinants of gap, model and market spreads are, to some extent, time variant, whereas CRI and equity returns are always statistically significant with comparable signs. Especially the slope of CRI and the credit curve contribute over-proportional to the explanatory value in times of financial turmoil, indicating the importance of market sentiment on the credit protection of the buyer and seller side. I can show that illiquidity spillover effects from equity to credit markets exist for market spreads in the financial crisis, highlighting the importance of detecting time variation in determinants. My analysis can be expanded in future research based on two main aspects. A first extension might be related to the model calibration itself. Better data availability with access to proprietary databases such as Markit, Moodys or S&P could yield more accurate recovery rates of underlying bonds, thus most likely improving model performance. Additionally, different asset volatility approximations could be used such as EWMA-, GARCH models or different historical volatility windows to derive a more granular view on the calibration contingent model performance outlined by my results. A second extension might focus on my regression approach. The frequency of weekly observations could be adjusted to match observation frequencies for which more macroeconomic variables are reported, e.g. monthly, quarterly or bi-annually. This would allow factors like GDP or unemployment rates to enter the regression specification and could provide additional insights. To compensate the number of reduced data points, a longer observation period can be defined. This is, however, only possible if other datasources with long back reaching historical data are available. Furthermore, different econometric techniques can be used and compared, such as panel quantile regressions or Markov regime switching models in style of Alexander and Kaeck (2008). With these methods it would be possible to correct results for path dependencies induced by selected econometric methods, identifying regression model inherent tendencies. In addition, further liquidity measures, such as the bid-ask spread, could be used to further assess the role of possible illiquidity in credit derivatives markets. Finally, fu-

ture research could test the above mentioned aspects in performing capital structure arbitrage strategies where different model specifications and subsequent risk-return profiles are compared.

References

Alexander, C., & Kaeck, A. (2008). Regime dependent determinants of credit default swap spreads. *Journal of Banking & Finance*, *32*(6), 1008–1021.

Allen, F., & Santomero, A. M. (1997). The Theory of Financial Intermediation. *Journal of Banking & Finance*, *21*(11-12), 1461–1485.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, *5*(1), 31–56.

Anderson, R., & Sundaresan, S. (2000). A comparative study of structural models of corporate bond yields: An exploratory investigation. *Journal of Banking & Finance*, *24*(1-2), 255–269.

Arora, N., Gandhi, P., & Longstaff, F. A. (2012). Counterparty credit risk and the credit default swap market. *Journal of Financial Economics*, *103*(2), 280–293.

Aunon-Nerin, D., Cossin, D., Hricko, T., & Huang, Z. (2002). Exploring for the determinants of credit risk in credit default swap transaction data: Is fixed-income markets' information sufficient to evaluate credit risk? *FAME Research Paper*(65).

Bajlum, C., & Larsen, P. T. (2008). *Capital Structure Arbitrage: Model Choice and Volatility Calibration* (Working Paper). Copenhagen Business School and Denmarks National Bank.

Bedendo, M., Cathcart, L., & El-Jahel, L. (2011). Market and model credit default swap spreads: mind the gap! *European Financial Management*, *17*(4), 655–678.

Bedendo, M., Cathcart, L., El-Jahel, L., & Liesch, L. (2005). Trading down the slope(s). *Risk Magazine*, *18*, 107–109.

Benkert, C. (2004). Explaining credit default swap premia. *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, *24*(1), 71–92.

Berndt, A., Jarrow, R. A., & Kang, C. (2007). Restructuring Risk in Credit Default Swaps: An Empirical Analysis. *Stochastic Processes and their Applications*, *117*(11), 1724–1749.

BIS. (2000). *Principles for the Management of Credit Risk* (White Paper). Bank for International Settlements - Basel Committee on Banking Supervision.

Black, F., & Cox, J. C. (1976). Valuing Corporate Securities: Some Effects of Bond Indenture Provisions. *The Journal of Finance*, *31*(2), 351–367.

Black, F., & Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, *81*(3), 637–654.

Blanco, R., Brennan, S., & Marsh, I. (2003). An Emprical Analysis of the Dynamic Relationship between Investment Grade Bonds and Default Swaps. *Bank of England*.

Bongaerts, D., De Jong, F., & Driessen, J. (2011). Derivative pricing with liquidity risk: Theory and evidence from the credit default swap market. *The Journal of Finance*, *66*(1), 203–240.

Bongaerts, D., De Jong, F., & Driessen, J. (2017). An asset pricing approach to liquidity effects in corporate bond markets. *The Review of Financial Studies*, *30*(4), 1229–1269.

Boyarchenko, S. (2000). *Endogenous default under Lévy processes* (Tech. Rep.). Working Paper, University of Texas, Austin.

Brandon, K., & Fernandez, F. (2005). Financial Innovation and Risk Management: An Introduction to Credit Derivatives. *Journal of Applied Finance*, *15*(1).

Byström, H. (2006). CreditGrades and the iTraxx CDS index market. *Financial Analysts Journal*, *62*(6), 65–76.

Campbell, J. Y., Lettau, M., Malkiel, B. G., & Xu, Y. (2001). Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk. *The Journal of Finance*, *56*(1), 1–43.

Campbell, J. Y., & Taksler, G. B. (2003). Equity volatility and corporate bond yields. *The Journal of Finance*, *58*(6), 2321–2350.

Cao, C., Yu, F., & Zhong, Z. (2010). The information content of option-implied volatility for credit default swap valuation. *Journal of Financial Markets*, *13*(3), 321–343.

Cao, C., Yu, F., & Zhong, Z. (2011). Pricing credit default swaps with option-implied volatility. *Financial Analysts Journal*, *67*(4), 67–76.

Chan, R. H., Guo, Y. Z. Y., Lee, S. T., & Li, X. (2019). *Financial Mathematics, Derivatives and Structured Products*. Springer Singapore.

Chen, R.-R., Fabozzi, F. J., & Sverdlove, R. (2010). Corporate Credit Default Swap Liquidity and Its Implicationsfor Corporate Bond Spreads. *The Journal of Fixed Income*, *20*(2), 31–57.

Collin-Dufresne, P., & Goldstein, R. S. (2001). Do Credit Spreads Reflect Stationary Leverage Ratios? *The Journal of Finance*, *56*(5), 1929–1957.

Collin-Dufresne, P., Goldstein, R. S., & Martin, J. S. (2001). The determinants of credit spread changes. *The Journal of Finance*, *56*(6), 2177–2207.

Corey, D. M., Dunlap, W. P., & Burke, M. J. (1998). Averaging correlations: Expected values and bias in combined Pearson rs and Fisher's z transformations. *The Journal of General Psychology*, *125*(3), 245–261.

Cox, J. C., Ingersoll Jr, J. E., & Ross, S. A. (1985). A Theory of the Term Structure of Interest Rates. *Econometrica*, *53*(2), 385–408.

Cremers, M., Driessen, J., Maenhout, P., & Weinbaum, D. (2008). Individual stock-option prices and credit spreads. *Journal of Banking & Finance*, *32*(12), 2706–2715.

Das, S. R., & Sundaram, R. K. (2001). A Discrete–Time Approach To Arbitrage-Free Pricing Of Credit Derivatives. In *Quantitative Analysis In Financial Markets: Collected Papers of the New York University Mathematical Finance Seminar (Volume III)* (pp. 86–109).

De Jong, F., & Driessen, J. (2012). Liquidity risk premia in corporate bond markets. *The Quarterly Journal of Finance*, *2*(2).

Diamond, D. W. (1984). Financial Intermediation and Delegated Monitoring. *The Review of Economic Studies*, *51*(3), 393–414.

Di Cesare, A., & Guazzarotti, G. (2010). An analysis of the determinants of credit default swap spread changes before and during the subprime financial turmoil. *Bank of Italy Temi di Discussione (Working Paper)*, *749*.

Driscoll, J. C., & Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics*, *80*(4), 549–560.

Duarte, J., Longstaff, F. A., & Yu, F. (2007). Risk and Return in Fixed-Income Arbitrage: Nickels in Front of a Steamroller? *The Review of Financial Studies*, *20*(3), 769–811.

Duffee, G. R. (1998). The relation between treasury yields and corporate bond yield spreads. *The Journal of Finance*, *53*(6), 2225–2241.

Duffie, D., & Singleton, K. J. (1999). Modeling Term Structures of Defaultable Bonds. *The Review of Financial Studies*, *12*(4), 687–720.

Duffie, D., & Singleton, K. J. (2003). *Credit risk: pricing, measurement, and management*. Princeton University Press.

Eom, Y. H., Helwege, J., & Huang, J.-z. (2004). Structural models of corporate bond pricing: An empirical analysis. *The Review of Financial Studies*, *17*(2), 499–544.

Ericsson, J., Jacobs, K., & Oviedo, R. (2009). The determinants of credit default swap premia. *Journal of Financial and Quantitative Analysis*, *44*(1), 109–132.

Ericsson, J., & Reneby, J. (2001). *The valuation of corporate liabilities: Theory and tests* (Tech. Rep.). SSE/EFI Working Paper Series in Economics and Finance.

Ericsson, J., Reneby, J., & Wang, H. (2006). *Can structural models price default risk? Evidence from bond and credit derivative markets* (Tech. Rep.). Citeseer.

Finger, C. C., Finkelstein, V., Lardy, J.-P., Pan, G., Ta, T., & Tierney, J. (2002, May). *CreditGrades* (Tech. Rep.). RiskMetrics Group Inc., New York.

Gambetti, P., Gauthier, G., & Vrins, F. (2019). Recovery rates: Uncertainty certainly matters. *Journal of Banking & Finance*, *106*, 371–383.

Geske, R. (1977). The Valuation of Corporate Liabilities as Compound Options. *The Journal of Financial and Quantitative Analysis*, *12*(4), 541–552.

Goyal, A., & Santa-Clara, P. (2003). Idiosyncratic risk matters! *The Journal of Finance*, *58*(3), 975–1007.

Hilberink, B., & Rogers, L. C. (2002). Optimal capital structure and endogenous default. *Finance and Stochastics*, *6*(2), 237–263.

Hogan, S., Jarrow, R., Teo, M., & Warachka, M. (2004). Testing Market Efficiency using Statistical Arbitrage with Applications to Momentum and Value Strategies. *Journal of Financial Economics*, *73*(3), 525–565.

Houweling, P., & Vorst, T. (2005). Pricing default swaps: Empirical evidence. *Journal of International Money and Finance*, *24*(8), 1200–1225.

Hu, H., & Lawrence, L. (2000). Estimating recovery rates. *Internal document, JPMorgan*.

Huang, Z. J., & Luo, Y. (2016). Revisiting Structural Modeling of Credit Risk—Evidence from the Credit Default Swap (CDS) Market. *Journal*

of Risk and Financial Management, *9*(2), 3.

Hull, J. (2012). *Options, Futures, and other Derivatives* (8th ed.). Boston, Prentice Hall.

Hull, J., Predescu, M., & White, A. (2004). The relationship between credit default swap spreads, bond yields, and credit rating announcements. *Journal of Banking & Finance*, *28*(11), 2789–2811.

Hull, J., & White, A. (2000). Valuing Credit Default Swaps I: No Counterparty Default Risk. *The Journal of Derivatives*, *8*(1), 29–40.

Hunter, J. E., & Schmidt, F. L. (1990). Dichotomization of continuous variables: The implications for meta-analysis. *Journal of Applied Psychology*, *75*(3), 334–349.

Imbierowicz, B., & Cserna, B. (2008). How Efficient are Credit Default Swap Markets? An Empirical Study of Capital Structure Arbitrage based on Structural Pricing Models. In *21st Australasian Finance and Banking Conference.*

Jarrow, R., & Protter, P. (2004). Structural versus Reduced Form Models: a new information based perspective. *Journal of Investment Management*, *2*(2), 1–10.

Jarrow, R. A., & Turnbull, S. M. (1995). Pricing Derivatives on Financial Securities subject to Credit Risk. *The Journal of Finance*, *50*(1), 53–85.

Jones, E. P., Mason, S. P., & Rosenfeld, E. (1984). Contingent claims analysis of corporate capital structures: An empirical investigation. *The Journal of Finance*, *39*(3), 611–625.

Junge, B., & Trolle, A. B. (2015). Liquidity risk in credit default swap markets. *Swiss Finance Institute Research Paper*(13-65).

Kiesel, R., & Veraart, L. A. (2008). A note on the survival probability in CreditGrades. *Journal of Credit Risk*, *4*(2), 65–75.

Kim, I. J., Ramaswamy, K., & Sundaresan, S. (1993). Does default risk in coupons affect the valuation of corporate bonds?: A contingent claims model. *Journal of Financial Management*, *22*(3), 117–131.

Klieštik, T., & Cúg, J. (2015). Comparison of Selected Models of Credit Risk. *Procedia Economics and Finance*, *23*, 356–361.

Lando, D. (2009). *Credit risk modeling: theory and applications*. Princeton University Press.

Leland, H. E. (1994). Corporate debt value, bond covenants, and optimal capital structure. *The Journal of Finance*, *49*(4), 1213–1252.

Leland, H. E., & Toft, K. B. (1996). Optimal Capital Structure, Endogenous Bankruptcy, and the Term Structure of Credit Spreads. *The Journal of Finance*, *51*(3), 987–1019.

Levin, A., Lin, C.-F., & Chu, C.-S. J. (2002). Unit root tests in panel data: Asymptotic and finite-sample properties. *Journal of Econometrics*, *108*(1), 1–24.

Longstaff, F. A., Mithal, S., & Neis, E. (2005). Corporate yield spreads: Default risk or liquidity? New evidence from the credit default swap market. *The Journal of Finance*, *60*(5), 2213–2253.

Longstaff, F. A., & Schwartz, E. S. (1995). A simple approach to valuing risky fixed and floating rate debt. *The Journal of Finance*, *50*(3), 789–819.

Madan, D., & Unal, H. (2000). A Two-Factor Hazard Rate Model for Pricing Risky Debt and the Term Structure of Credit Spreads. *Journal of Financial and Quantitative Analysis*, *35*(1), 43–65.

Merton, R. C. (1974). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *The Journal of Finance*, *29*(2), 449–470.

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

Musiela, M., & Rutkowski, M. (1998). *Martingale Methods in Financial Markets*. Springer.

Norden, L., & Weber, M. (2009). The co-movement of credit default swap, bond and stock markets: An empirical analysis. *European Financial Management*, *15*(3), 529–562.

Pástor, L., & Stambaugh, R. F. (2003). Liquidity risk and expected stock returns. *Journal of Political Economy*, *111*(3), 642–685.

Penza, P., Bansal, V. K., Bansal, V. K., & Bansal, V. K. (2001). *Measuring market risk with value at risk* (Vol. 17). John Wiley & Sons.

Remolona, E. M., Wooldridge, P. D., et al. (2003). The euro interest rate swap market. *BIS Quarterly Review*, *47*.

Rubinstein, M., & Reiner, E. (1991). Breaking down the barriers. *Risk*, *4*, 28–35.

Skinner, F. S., & Townend, T. G. (2002). An empirical analysis of credit default swaps. *International Review of Financial Analysis*, *11*(3), 297–309.

Stamicar, R., & Finger, C. (2006). Incorporating equity derivatives into the CreditGrades model. *Journal of Credit Risk*, *2*(1), 1–20.

Tang, D. Y., & Yan, H. (2007). *Liquidity and credit default swap spreads.* https://ssrn.com/abstract=1008325. (Available at SSRN 1008325)

Thomas, C. W. (2002). The rise and fall of Enron. *Journal of Accountancy*, *193*(4), 41–52.

Tsuji, C. (2005). The credit-spread puzzle. *Journal of International Money and Finance*, *24*(7), 1073–1089.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, *5*(2), 177–188.

Wagner, N. (2008). *Credit risk: Models, derivatives, and management* (Financial Mathematics Series, 12. ed.). CRC press - Chapman & Hall.

Wojtowicz, M. (2014). *Capital Structure Arbitrage revisited* (Tech. Rep.). Amsterdam & Rotterdam, Tinbergen Institute.

Yu, F. (2006). How Profitable is Capital Structure Arbitrage? *Financial Analysts Journal*, *62*(5), 47–62.

Zhang, B. Y., Zhou, H., & Zhu, H. (2009). Explaining credit default swap spreads with the equity volatility and jump risks of individual firms. *The Review of Financial Studies*, *22*(12), 5099–5131.

Zhou, C. (2001). The Term Structure of Credit Spreads with Jump Risk. *Journal of Banking & Finance*, *25*(11), 2015–2040.

Junior Management Science 8(1) (2023) 43-67

Junior Management Science

journal homepage: www.jums.academy

From investor to entrepreneur – An explorative study of the entrepreneurial behaviour of investor-entrepreneurs

Moritz Setzer

Technische Universität München

Abstract

Research on entrepreneurial types argues that differences in decisions and actions of novice and habitual entrepreneurs can be attributed to prior experience and social ties. However, other forms of resource acquisition than business ownership experience are neglected in previous studies. This study investigates this research gap by exploring the entrepreneurial behavior of novice investor-entrepreneurs, first-time founders who have previously worked in venture capital or private equity. Drawing on a cross-sectional approach with 13 semi-structured interviews, this paper investigates whether investor-entrepreneurs have already acquired the necessary resources relevant for new venture foundations in their role as investors. This study reveals that investor-entrepreneurs differ from other first-time entrepreneurs in several aspects. It shows that the differences can mainly be attributed to skills and knowledge already acquired through the job as an investor. In fact, venture capital experience seems to be more helpful for a venture's early foundation than private equity experience. The study discusses the implications of these findings for the literature on entrepreneurial types and their differentiation, thereby challenging existing differentiation and classification approaches.

Keywords: Investor-entrepreneur; Entrepreneurial types; Novice; Habitual.

1. Introduction

Shane and Venkataraman (2000, p. 218) define entrepreneurship as a multi-layered framework comprising "sources of opportunities; the processes of discovering, evaluating and exploiting opportunities; and the group of individuals who discover, evaluate and exploit them". In line with this definition, the entrepreneurial literature traditionally has focused on examining successful founders, their characteristics, processes, and sources of information, in order to define successful entrepreneurship. MacMillan (1986, p. 242), in his ground-breaking call "To really learn about entrepreneurship, let's study habitual entrepreneurs" distinguishes between novice and habitual founders based on previous business ownership experience. He argues that habitual founders learn the necessary skills and knowledge for successful entrepreneurship from previous business foundations (MacMillan, 1986). Thus, by studying habitual entrepreneurs, essential skills and knowledge that make the entrepreneurial process more efficient could be identified (MacMillan, 1986). Following this invocation, multiple studies have taken up this pursuit and conducted generic comparisons between these two types of entrepreneurs, asking different research questions and using different units of analysis, theoretical perspectives and methods (Birley & Westhead, 1993, pp. 40-41; Donckels, Dupont, & Michel, 1987, pp. 48-49; Kolvereid & Bullvag, 1993, pp. 276-278; Mosey & Wright, 2007, pp. 911-917; Politis, 2008, pp. 476-479). Impelled by MacMillan (1986) focus on habitual founders, more refined studies followed, researching differences in performance and characteristics along habitual sub-groups (Westhead, Ucbasaran, & Wright, 2005a, 2005b, pp. 396-398; Westhead, Ucbasaran, & Wright, 2005b, pp. 75-77; Westhead & Wright, 1998a, p. 176). At the same time, respective studies continuously consider novice entrepreneurs as a homogeneous group of first-time founders as the basis for comparison (e.g., Westhead & Wright, 1998a, p. 176).

Yet, none of these studies finds that habitual founders run more successful businesses than inexperienced novice counterparts (e.g., Kolvereid & Bullvag, 1993, p. 280; Westhead & Wright, 1998a, p. 192). Following, scholars switch focus to founder-related differences between novice and habitual entrepreneurs rather than just exploring venture-level performance indicators. Today, novice and habitual entrepreneurs can be distinguished along their personal

DOI: https://doi.org/10.5282/jums/v8i1pp43-67

(e.g., age, gender, education) and entrepreneurial characteristics consisting of individual decisions and actions (e.g., opportunity identification, venture financing, motivation) (Birley & Westhead, 1993; Donckels et al., 1987; Kolvereid & Bullvag, 1993, pp. 278-281; Politis, 2008, pp. 480-485; Westhead, Ucbasaran, & Wright, 2005a, pp. 403-412; Westhead, Ucbasaran, & Wright, 2005b, pp. 86-89). Furthermore, differences in decisions and actions are attributed to differences in skills and knowledge (human capital) and social ties (social capital) (Mosey & Wright, 2007, p. 932; Westhead, Ucbasaran, & Wright, 2005a, pp. 413-414; Westhead, Ucbasaran, & Wright, 2005b, pp. 90-91). In line with the distinction of entrepreneurs according to prior business ownership experience, the stock of prior business ownership experience is also used as a proxy for the availability of entrepreneurial characteristics (e.g., Mosey & Wright, 2007, p. 912; Westhead, Ucbasaran, & Wright, 2005a, pp. 396-397; Westhead & Wright, 1998a, pp. 179-183).

But what about first-time founders who might have acquired the human and social capital necessary for a successful foundation other than through previous business ownership experience? Human capital theory suggests that HC useful for founding must be both process- and content-specific, not just the sum of experiences (Unger, Rauch, Frese, & Rosenbusch, 2011, p. 12; West & Noel, 2009, pp. 6-7). Thus, previous jobs with parallels to entrepreneurial tasks and a similar business model could compensate for the lack of previous entrepreneurial experience. Here, private equity investors might be a good example. By working in a venture capital (VC) or private equity (PE) fund and accompanying entrepreneurial tasks such as fundraising by sitting on the other side of the table, they might acquire superior task-related skills and knowledge compared to their "normal" novice counterparts. Furthermore, as they work together with many entrepreneurs, they may have learned useful characteristics from other entrepreneurs and build social ties that might be helpful for their own subsequent foundations. Thus, the following research question guides this study: *How do investor-entrepreneurs differ from novice and habitual entrepreneurs?* The aim is to compare the characteristics with what we know about established types (i.e., novice and habitual) and to identify the behaviour regarding entrepreneurial processes such as opportunity identification, network access, or knowledge resources. Furthermore, the goal is to integrate the investor-entrepreneur into the current theoretical perspective.

I conducted an explorative (qualitative) cross-sectional study with founders who have worked as investors in a venture capital or private equity firm before their foundation to answer the respective research question. To do so, I used 13 semi-structured interviews as the primary source of data that I conducted with relevant founders in Germany over a period of six weeks. My findings show that investor-entrepreneurs differ in several aspects from "normal" novice entrepreneurs with other backgrounds, investigated in prior studies, and therefore cannot be equated with other first-time founders. I also find that contrary to the original assumption, private equity investors (consisting of VC and PE investors) cannot be seen as a homogeneous group either. While they overall show differences, but also various similarities to "normal" novice entrepreneurs, particularly investor-entrepreneurs with previous VC experience differ in many ways such significantly in their respective characteristics that they cannot be assigned to existing types of entrepreneurs. Investor-entrepreneurs also tend to have operational and mindset shortcomings that help understand their entrepreneurial characteristics better. In a broader sense, shortcomings also represent a new influencing variable that previous studies have not considered. To put the investor-entrepreneur in a current theoretical perspective, first, this study calls for a more refined distinction of novice entrepreneurs on another level when applying the established classification based on prior business ownership experience. Second, the results of this study suggest a novel approach of differentiation based on a resource-based view. Thereby, it indicates that investor-entrepreneurs have a more advanced “technology of entrepreneurship” (MacMillan, 1986) compared to other first-time founders. Lastly, I also provide novel insights on the founder’s opportunity identification and evaluation process.

These findings contribute to our knowledge of entrepreneurial types, their distinction and entrepreneurship more broadly. First, previous research (e.g., Kolvereid & Bullvag, 1993; Mosey & Wright, 2007; Westhead, Ucbasaran, Wright, & Binks, 2005) has focused on classifying entrepreneurs based on previous business ownership experience to compare them. In contrast to a homogeneous view of novice entrepreneurs, this study shows that this cannot be the case. Furthermore, I suggest that the measurement of entrepreneurial skills and knowledge should no longer be on the stock of experience (e.g., Westhead & Wright, 1998a, pp. 179-183), but preferably on the task-relatedness of experience. The sum of implications extends existing models on intervariable linkages of distinguishing variables. Second, my side findings on entrepreneurial behaviour have broader implications on the entrepreneurial literature regarding opportunity perception (e.g., Baron & Ensley, 2006; Shane & Venkataraman, 2000; Venkataraman, 1997) by providing a new influencing variable in opportunity identification and novel insights into founders’ physical evaluation processes.

This paper is structured as follows: The theoretical and conceptual framework in chapter 2 aims to create a coherent knowledge base regarding key concepts and existing literature, forming the basis for analysing the investor-entrepreneur. Chapter 3 discusses the data collection and research methodology. Chapter 4 includes the presentation and discussion of the results. Here I also put my findings in relation to existing literature. The aim of Chapter 5 is to put the results on a deeper level of discussion, elaborate implications for academics and practitioners and discusses the limitations of this study.

2. Theoretical background

I base this study on two lines of literature. The first comprises entrepreneurship, i.e., the distinction between different types of entrepreneurs and the factors influencing their entrepreneurial behaviour and performance. The second focuses on private equity as well as its stakeholders and their job-specific characteristics. In the following chapter, I examine the necessary literature, review it critically and derive a necessary theoretical framework.

2.1. Types of entrepreneurs

For a long time now, entrepreneurial literature has distinguished between different types of entrepreneurs. Well known and frequently referred to examples are Smith (1967) with his definition of the craftsman and opportunistic entrepreneur, Schumpeter (1934), who distinguished the equilibrium-breaking entrepreneur and Kirzner (1973, pp. 72-73) with his antithetic equilibrium-restoring entrepreneur. The lack of consistent and pertinent definitions between different types of entrepreneurs primarily results from different perceptions of entrepreneurship due to a lack of general definition (Shane & Venkataraman, 2000, p. 217). Nevertheless, an established categorization in the entrepreneurial literature is the distinction between novice and habitual entrepreneurs, primarily based on previous business ownership experience (Birley & Westhead, 1993, p. 40; MacMillan, 1986; Mosey & Wright, 2007; Politis, 2008, p. 475). *Novice* founders thereby describe founders without prior minority or majority business ownership experience but who have founded or acquired a company and own it at the time of observation (e.g., Birley & Westhead, 1993, p. 40; Westhead & Wright, 1998a, p. 176). On the other hand, *habitual* entrepreneurs are founders with previous business experience who have started or owned at least one business prior to the current one (e.g., Westhead & Wright, 1998a, p. 176).

Based on MacMillan (1986) call to study habitual entrepreneurs in greater depth, various researchers have focused their studies on different types of repeated entrepreneurs. Despite extensive research since then, yet there is no unified definition of the term "habitual". Kolvereid and Bullvag (1993, p. 276) or Birley and Westhead (1993, p. 40), for example, define founders who have founded at least one company other than the current one as "habitual" founders. Other studies show that you cannot see them as a homogeneous group and distinguish respective entrepreneurs at a second level of depth, based on their prior venture exit. Here, entrepreneurs who have sold or exited their previous company before starting their current one are called "serial" founders. Entrepreneurs who continue to own all or parts of their previous company are called "portfolio" founders (P. Hall, 1995, p. 220; Westhead & Wright, 1998b, p. 65). Based on the delineation characteristics described above, it is evident that novice entrepreneurs who are repeatedly involved in new venture creation become habitual entrepreneurs (Westhead & Wright, 1998a, p. 176). Figure 1 summarises the classification into different types of entrepreneurs.

In contrast to the literature's extensive differentiation and consideration of habitual entrepreneurs, novice entrepreneurs are not further differentiated and viewed as a homogeneous group. For example, Westhead, Ucbasaran, and Wright (2005b, p. 77) cite lack of start-up experience as a consistent disadvantage of novice entrepreneurs, which causes them to be "overwhelmed by information and/or not know how to use the information". Kolvereid and Bullvag (1993, p. 275) use the words "average" and "typical" to describe novice entrepreneurs, thus highlighting a homogeneous perspective. Consequently, the distinction of habitual founders established in the 1990s is maintained in recent studies, while novice entrepreneurs are still considered homogeneous in more recent studies (e.g., Carbonara, Tran, & Santarelli, 2019, p. 123; Plehn-Dujowich, 2010, p. 377; St-Jean & Audet, 2012). Robbie and Wright (1996, pp. 71-81), with their distinction between buy-out and buy-in entrepreneurs, are in fact among the only researchers to consider differences in terms of character traits among initial entrepreneurs, even though they do not use the term "novice".

2.2. Entrepreneurial differentiation variables

In line with the classification of different types of entrepreneurs, researchers' interest regarding possible differentiating variables has increased. Donckels et al. (1987), for example, studied 400 repeated (i.e., habitual) and "other" founders in Belgium concerning their personal profile, strategic behaviour and characteristics. Kolvereid and Bullvag (1993) examine 250 Norwegian novice and habitual founders along multiple personal characteristics and test their findings on British entrepreneurs. In several studies, a group of authors around Paul Westhead examine founders from Great Britain (sometimes only Scotland) along various characteristics such as family background, work experience, skills, knowledge, cognitive abilities, motivation, and their business performance (Birley & Westhead, 1993; Westhead, Ucbasaran, & Wright, 2005a, Westhead, Ucbasaran, Wright, & Binks, 2005; Westhead & Wright, 1998b). Politis (2008) investigates whether previous business ownership experience impacts entrepreneurial characteristics (e.g., coping with liabilities of newness) using 231 entrepreneurs in Sweden. Zhang (2011) focuses on a specific entrepreneurial process, namely venture capital acquisition. Drawing upon statistical analysis of 10,000 entrepreneurs, she examines whether prior business ownership experience has impacts time and amount of venture capital acquisition. Mosey and Wright (2007) investigate 24 academic entrepreneurs and studied their ability to develop social contacts as well as their network usage.

Thus far, none of these studies finds evidence that habitual founders operate more successful businesses than inexperienced novice counterparts (e.g., Kolvereid & Bullvag, 1993, p. 280; Westhead & Wright, 1998a, p. 192). One possible explanation for this recurring finding is that

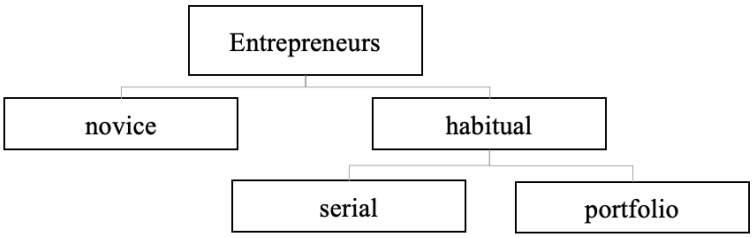

Figure 1: Types of entrepreneurs.

most studies use statistical tests of limited variables to find significant correlations of means between novice and habitual entrepreneurs venture performance while applying varying definitions of the individual founders (especially for habitual founders) and different definitions of performance and success factors. Moreover, in addition to the type of entrepreneur, other factors such as geographical location, industry or strategic market positioning influence ventures' performance (e.g., Sandberg & Hofer, 1987, pp. 16-21). Consequently, scholars switch focus to founder-related differences between novice and habitual entrepreneurs despite only searching for venture-level performance indicators. Today, novice and habitual entrepreneurs can be distinguished along their personal (e.g. age, gender, education) and entrepreneurial characteristics, including their decisions and actions regarding essential tasks (e.g., opportunity identification, venture financing) (Birley & Westhead, 1993, pp. 41-53; Kolvereid & Bullvag, 1993, pp. 278-281; Politis, 2008, pp. 480-485; Westhead, Ucbasaran, & Wright, 2005a, pp. 403-412; Westhead, Ucbasaran, & Wright, 2005b, pp. 86-89; Westhead, Ucbasaran, Wright, & Binks, 2005, pp. 114-126). Furthermore, differences in decisions and actions are attributed to differences in stock of experience and social ties (Mosey & Wright, 2007, p. 930; Westhead, Ucbasaran, & Wright, 2005a, pp. 412-413; Westhead, Ucbasaran, & Wright, 2005b, pp. 89-91).

2.2.1. Human Capital

Becker (1964, pp. 15-17) defines *human capital* as the sum of knowledge and skills individuals can acquire through investments in education, training, and other forms of experience. Becker's concept further distinguishes between human capital investments and human capital investment outcomes, as well as task-related HC and non-task-related HC (Becker, 1964, pp. 16-18). Human capital investments thereby comprise the actual experiences, i.e., training and work experience, whereas human capital outcomes describe the skills and knowledge that may arise through experiences (Unger et al., 2011, p. 7). Task-relatedness refers to whether the investment in human capital and the results are linked to a specific task, such as operating a venture (Unger et al., 2011, p. 11). Human capital (HC) represents an important field of research in the entrepreneurial literature, as there is an overall positive relationship between human capital and the success of new ventures (Unger et al., 2011, p. 23). Furthermore, Unger et al. (2011, p. 23) shows that the correlation of task-related HC and HC investment outcomes is higher than for their composites.

Resource-based theory perceives knowledge and skills (i.e., human capital outcomes) as a resource that influences entrepreneurial capabilities and ultimately a firm's competitive advantage (Grant, 1991, p. 115; West & Noel, 2009, pp. 1-2). In the process of starting a business, knowledge (i.e., resources) about the particular industry, the type of business and the process of starting a business is needed (Wiklund & Shepherd, 2003, pp. 1308-1310). Thereby, the respective knowledge influences essential entrepreneurial capabilities such as opportunity identification and exploitation, strategy formulation and planning (Chandler & Jansen, 1992, p. 225; Shane & Venkataraman, 2000, pp. 220-224). Thus, differences in entrepreneurial behaviour and characteristics can be attributed to differences in human capital.

To measure HC, or knowledge and skills more definite, experience can either be seen as "stock" or "stream" (Reuber & Fischer, 1999). Thereby, stock refers to the sum of an individual's experiences at a certain point in time. In contrast, when considered as stream, it refers to the entrepreneurial process of learning through experiencing, thus focusing on changes in what is usually done (Reuber & Fischer, 1999, p. 31). However, considering it as stock assumes a linear relationship between HC investments and HC investment outcomes, which may not be the case (Unger et al., 2011, p. 10; 26; West & Noel, 2009, p. 17). For example, school experience is often assumed to be a proxy for knowledge. Yet, what knowledge has really been acquired depends on the person himself and other circumstances such as the school characteristics and the genetic environment (Unger et al., 2011, p. 11). For this reason, stock of experience should not be considered the same as skills and knowledge (i.e., human capital). Measuring it as a stream emphasises the actual learning process by focusing on specific sequences of experiences. Yet, the problem with measuring experiences as a stream is that it is neither traceable nor easily comparable (Reuber & Fischer, 1999, p. 39). However, multiple studies use stock of business ownership experience as a proxy for important entrepreneurial HC (e.g., Politis, 2008, pp. 476-478; Westhead, Ucbasaran, & Wright, 2005a, p. 397; Zhang, 2011, p. 9).

Overall, differences in human capital exist, because (i) individuals have different initial resources (e.g., parental background, gender, education) which may influence the life ex-

perience and shape the basis for further HC investments and (ii) individuals expect compensation for their investments in their human capital and thus have different stocks of experience (Becker, 1964, pp. 85-89; Becker, 1993, p. 392). Differences in human capital between novice and habitual entrepreneurs can be identified across both grounds. (Cooper, Gimeno-Gascon, & Woo, 1994, p. 376) argue that initial resources such as gender and education (i.e., general human capital) influence an individual's life experience and thus contribute to the success of an entrepreneur. In general, novice and habitual founders are predominantly male, with habitual founders accounting for a larger share (Birley & Westhead, 1993, p. 44; Kolvereid & Bullvag, 1993, p. 279). While certain studies show that experienced founders have enjoyed higher education (Donckels et al., 1987, p. 52; Kolvereid & Bullvag, 1993, p. 278), others could not find significant differences (Birley & Westhead, 1993, p. 42; Westhead, Ucbasaran, Wright, & Binks, 2005, p. 114). This can be attributed to different perspectives on habitual entrepreneurs (habitual vs serial and portfolio) among researchers. The age of an entrepreneur is not inevitably decisive for the success of a venture but influences the knowledge of an entrepreneur (e.g., West & Noel, 2009, p. 6; Westhead, Ucbasaran, & Wright, 2005b, p. 74). Kolvereid and Bullvag (1993, pp. 278-279) found that habitual entrepreneurs start their first venture at a younger age.[1] In addition to supporting findings from other studies (Birley & Westhead, 1993, p. 42), Westhead and Wright report the same results in later studies with a refined differentiation of habitual entrepreneurs by serial and portfolio (Westhead & Wright, 1998b, p. 75; Westhead & Wright, 1998a, p. 186).

Regarding the stock of experience, habitual entrepreneurs tend to work for more organizations before becoming a founder themselves (Birley & Westhead, 1993, p. 43; Westhead, Ucbasaran, & Wright, 2005b, p. 79).[2] One explanation for this is that it is in the nature of the habitual entrepreneur to achieve a particular goal (e.g., successfully founding a company) and then seek a new challenge (MacMillan, 1986). In addition, novice entrepreneurs tend to have worked in a big company (>1.000 employees) before becoming a founder, while habitual are significantly more likely to have worked in a small company (<100 employees) (Birley & Westhead, 1993, p. 45).

2.2.2. Social capital

Social capital describes the existing and potential resources available from an individual's personal network of relationships (Stam, Arzlanian, & Elfring, 2014, p. 154). In this paper, a personal network defines an entrepreneur's family, friends and business contacts (e.g., colleagues, suppliers, customers) and their relations to each other (Dubini & Aldrich, 1991, p. 307; Witt, 2004, p. 392). Connections can be distinguished between strong and weak ties based on their strength which is influenced by emotional background, intimacy and amount of time spent (Granovetter, 1973, p. 1361). Unlike other forms of capital, SC is neither anchored in the actors themselves nor in their ties to each other (Coleman, 1988, p. 98), but rather is the outcome of investments in social relations and available to individuals as goodwill that can be used to achieve possible goals (Adler & Kwon, 2002, p. 18; 23).

Social capital matters in the entrepreneurial literature since prior studies find a positive correlation between social capital and entrepreneurial success (e.g., Stam et al., 2014, p. 163). Thereby, the personal network has an influence on the foundation ("network founding hypothesis") as well as on the subsequent organizational success ("network success hypothesis") of the new venture (Brüderl & Preisendörfer, 1998, p. 214). Thus, differences in entrepreneurial behaviour can be attributed to differences in social capital.

The literature argues that network connections enable an entrepreneur to identify new business opportunities (Ozgen & Baron, 2007, pp. 186-187), gain validation/legitimacy from external stakeholders in the entrepreneurial ecosystem (Elfring & Hulsink, 2003, p. 413) and grant them access to resources they would not be able to on the market (Witt, 2004, p. 392). Some examples: First, as mentioned earlier, information is important to identify opportunities. Further, Brüderl and Preisendörfer (1998, p. 214) argue that information from informal sources is more valuable, reliable and exclusive. Thus, a more extensive personal network including weak ties can provide more helpful information and thus more business opportunities. Second, entrepreneurs require different types of resources to be successful with their new venture, including physical, financial, and human resources (West & Noel, 2009, p. 1). However, due to the "liability of newness" and the associated information asymmetries between resource providers and new ventures at the beginning, resource providers are usually sceptical about the quality and legitimacy of the venture, which leads to the entrepreneurs lacking necessary resources (Shane, 2003, pp. 164-167). Strong ties and frequent interaction with one's network can help to overcome information asymmetries, as both the trust associated with strong ties and the finer exchange of information through regularity help to build legitimacy and exclude possible opportunistic behaviour by entrepreneurs (Uzzi, 1997, pp. 45-46). Third, Birley (1985, p. 113) shows that entrepreneurs mainly use their informal network, mostly business contacts, to acquire necessary resources for their business, e.g., the choice of location as well as the procurement of employees. In addition, research shows that entrepreneurs with greater social capital have superior access to financial resources (Shane & Cable, 2002, p. 374; Uzzi, 1999, p. 498; Zhang, 2011, p. 22). Fourth, because social capital provides access to resources, it can also be used to supplement missing or lacking HC (Aldrich & Zim-

[1] Kolvereid and Bullvag (1993): Age of founding first venture (mean): novice (35), habitual (30); Gender: novice: male (80%), habitual: male (97%); Birley and Westhead (1993): Age of founding first venture (mean): novice (36), habitual (30); Gender: novice: male (92%), habitual: male (94%).

[2] Novice entrepreneurs: 3 organisations before founding; habitual: 4 organisations before founding.

mer, 1986, p. 20; Birley, 1985, p. 115). Entrepreneurs can substitute industry, market or founding experience, which they do not have due to lack of experience, with knowledge gained through their network (e.g., founding partner) (West & Noel, 2009, p. 9).

Even though the influence of social capital on entrepreneurial tasks has been outlined in several studies, only a few researchers have studied differences in social capital between novice and habitual entrepreneurs. One possible reason for this is that compared to human capital, social capital it is less tangible and therefore harder to evaluate (Coleman, 1988, p. 100). Mosey and Wright (2007, p. 916) study is one of the few that examines and distinguishes nascent[3], novice and habitual entrepreneurs along their network structure, content, and governance. They present several characteristics in which novice and habitual entrepreneurs differ (Mosey & Wright, 2007, pp. 918-924). First, in terms of network structure, habitual founders typically have more network ties when starting their new business. Furthermore, experienced founders are keen to expand their network compared to their novice counterparts. However, both types of entrepreneurs focus on building relationships with large companies as well as with investors and smaller companies. Secondly, they differ in terms of network content. Habitual entrepreneurs typically seek equity and management knowledge to support their business growth. On the other hand, novice ones first try to engage in the next stage of business development, meaning they seek proof-of-concept financing and industry knowledge about potential customers and their unmet needs. Third, differences along network governance. Novice entrepreneurs build new ties mainly through networking events and broker activities where a third party introduces one another. Habitual entrepreneurs, on the other hand, are more strategic in choosing and establishing new contacts. For example, they use their social capital specifically to close resource gaps. They also strengthen existing relationships that could be helpful and replace less effective ones with more useful ones.

2.2.3. Differences in entrepreneurial characteristics and behaviour

The stock of experience (i.e., human capital) (Reuber & Fischer, 1999, p. 32) as well as the personal network (i.e., social capital) (Shane, 2003, p. 49) of a founder, have an impact on the decisions and actions of entrepreneurs. In this context, research shows that these variables have a particular impact on crucial initial steps in the entrepreneurial process, such as opportunity identification, strategy formulation and financing (e.g., Unger et al., 2011, p. 28; Westhead, Ucbasaran, & Wright, 2005a, pp. 403-412; Zhang, 2011, pp. 17-21). Reuber and Fischer (1999, p. 31) present the intermediary relationships of human capital in a conceptual model. Westhead, Ucbasaran, and Wright (2005a, p. 395) extended this model by distinguishing types of entrepreneurs as well as their respective decisions, actions, performance, and aspirations. Following this and taking into account studies on social capital and its influence on entrepreneurial processes (e.g., Ozgen & Baron, 2007, pp. 186-187; Shane & Cable, 2002, p. 377) and on differences between entrepreneurial types (Mosey & Wright, 2007, pp. 925-926), this model can be extended to include social capital influences. Figure 5 (Appendix A) shows the corresponding correlations. The following chapter describes the respective entrepreneurial processes and examines differences along these characteristics for novice and habitual entrepreneurs based on findings from prior studies.

Venkataraman (1997, p. 120) defines entrepreneurship as the transformation of opportunities into future goods and services. Therefore, identifying and seizing opportunities is one of the most critical entrepreneurial skills (Shane, 2003, p. 18; Venkataraman, 1997, pp. 120-121). In literature, two common reasons why opportunities repeatedly arise are market imperfections (Kirzner, 1973, pp. 9-11) and environmental changes (e.g., technological, political, social, regulatory) (Schumpeter, 1934, pp. 66-67; Shane, 2003, pp. 23-28). To identify opportunities, one needs (i) information regarding the respective environment and (ii) a set of skills and insights to process this information in order to recognize the respective opportunities (Shane, 2003, pp. 45-50; Venkataraman, 1997, p. 124). Individuals obtain the respective knowledge in three ways: First, through active search. By actively searching for relevant information, rather than through random behaviour, the chance of finding information that leads to the identification of an entrepreneurial opportunity increases (Shane, 2003, p. 48). Second, through *knowledge corridors*. These are specific channels through which the individual can acquire idiosyncratic knowledge and are influenced through prior experience (Shane & Venkataraman, 2000, p. 222; Venkataraman, 1997, p. 122). Third, entrepreneurs seek and obtain information about opportunities from external sources, with social networks being one way of accessing information (e.g., Ozgen & Baron, 2007, p. 186; Shane, 2003, p. 49). The structure of the individual network is thereby decisive for the quality, quantity and speed with which information is received (cf. Chapter 2.2.2) (Shane, 2003, p. 49). With regard to the necessary recognition skills, particularly cognitive conditions play an important role (Shane & Venkataraman, 2000, p. 222). They help explain why some people can interpret information and recognise an underlying opportunity, while others cannot, given the same amount of information and are influenced by individual circumstances, including everyday life, internal knowledge (through work experience) and social relationships (Venkataraman, 1997, p. 122). This individual characteristic is also called *opportunity alertness* (Kirzner, 1973, pp. 39-42). The better this alertness is, the less information an entrepreneur needs to connect the dots and identify the respective opportunity.

After an opportunity is discovered, the entrepreneur has to evaluate it and decide whether to exploit it or not (Shane,

[3]Nascent entrepreneurs are individuals considering starting their own business but have not done so yet (Mosey & Wright, 2007, p. 910).

2003, p. 61). In general, entrepreneurs exploit opportunities when they believe they will earn back more than they invest (Venkataraman, 1997, pp. 132-133). Thereby, the individual opportunity value is influenced by non-psychological (e.g., education, age, social position) and psychological (e.g., motivation, cognitive traits) factors (Shane, 2003, p. 62). Thereby, different social types, for example, see different value creation in opportunities (Gruber & MacMillan, 2017, p. 280), thus see different paybacks. While Darwinian entrepreneurs pursue a private, economic self-interest, communitarian entrepreneurs want to support and be supported by their social community. Missionary entrepreneurs, however, pursue a mission, thus want to advance a cause (Gruber & MacMillan, 2017, p. 278). This can be condensed as the motivation of individual entrepreneurs influencing the decision. Furthermore, studies show that opportunity-specific (e.g. rarity, competition, age) (Haynie, Shepherd, & McMullen, 2009, p. 353) and entrepreneur-specific characteristics (e.g. human capital, emotions) (Haynie et al., 2009, p. 354; Welpe, Spörrle, Grichnik, Michl, & Audretsch, 2012, p. 17) influence the opportunity evaluation. However, yet little is known about the intrapersonal process of opportunity evaluation (Welpe et al., 2012, p. 2).

Novice and habitual entrepreneurs can be distinguished along the listed influencing variables. Westhead, Ucbasaran, and Wright (2005a, pp. 403-412) show that, habitual entrepreneurs use more sources of information than novice ones. More than two thirds, regardless of entrepreneurial type, use information from customers, other business owners and their personal network (friends and family). They also show that novice entrepreneurs are also less likely to use consultants, local enterprise development agencies and governmental sources. In addition, habitual founders make more use of information from bankers, VCs and business angels. These differences support earlier findings from Kolvereid and Bullvag (1993, p. 282), who found habitual entrepreneurs to be more "resourceful" and indicate that habitual entrepreneurs use additional sources they gained through prior business experience. Second, Westhead, Ucbasaran, and Wright (2005a, pp. 406-407) show that no significant difference exists between the acquisition of information (active search, knowledge corridor & social networks) and the types of entrepreneurs exist. Third, it is assumed that prior business ownership experience contributes positively to developing the knowledge corridor and the opportunity alertness (Westhead, Ucbasaran, & Wright, 2005a, p. 397). Thus, habitual entrepreneurs should identify more opportunities, what current research suggests is the case. Westhead, Ucbasaran, and Wright (2005a, p. 409) show that more novice entrepreneurs report having identified zero opportunities, while a significantly higher number of habitual entrepreneurs identified two or more opportunities, indicating higher opportunity alertness of habitual entrepreneurs. Indeed, more habitual than novice entrepreneurs agree with corresponding assessment affirmations[4] and the direct statement "I have a special alertness or sensitivity for spotting opportunities" (Westhead, Ucbasaran, & Wright, 2005a, p. 408). Fourth, concerning the exploitation of opportunities, differences in the individual motivation of entrepreneurs were researched. Westhead and Wright (1998a, p. 187), for example, show that "need for independence", "need for approval", and "perceive instrumentality of wealth" are the three most relatable motivations of the respective founders. The latter is mentioned more often by habitual founders, while novice founders more frequently mention the first two statements (Westhead & Wright, 1998a, p. 189). Findings by Donckels et al. (1987, pp. 53-54) suggest that new founders are driven by "job satisfaction", "personal independence", and the pursuit of "quality products", while habitual founders show similar characteristics but primarily seek "self-achievement". Both studies suggest a Darwinian social identity for both groups, putting themselves at the centre of interest and adhering to conventional business logic (Gruber & MacMillan, 2017, pp. 277-278).

Further behavioural distinctions can be found in the number of equity partners, the industry and the financing of the respective new business, depending on the type of entrepreneur. First, business partners can reduce risk and provide resources such as capital or management skills (Westhead, Ucbasaran, & Wright, 2005b, p. 76; 91); thus, most novice and habitual entrepreneurs tend not to start up their ventures alone (Birley & Westhead, 1993, p. 43; Kolvereid & Bullvag, 1993, p. 280; Westhead, Ucbasaran, & Wright, 2005b, p. 79). Second, Cooper (1985, pp. 84-85) shows that the employer can be an incubator for subsequent foundations with respect to geographical location, the nature of business, type of company (e.g., industrial firm, hospital or university) and the size of the respective venture. In line with this “incubator phenomenon” (Cooper, 1985, p. 84), more than one-third of each type of entrepreneur starts their business in the same industry as their last employer, whereby novice entrepreneurs are significantly more likely to do so (Birley & Westhead, 1993, p. 44; Politis, 2008, p. 479). Third, obtaining financial capital is an essential and challenging part of the entrepreneurial process (e.g., Birley, 1985, p. 115; West & Noel, 2009, p. 1; Westhead, Ucbasaran, & Wright, 2005a, p. 406). Personal savings and family and friends are the most frequently used source to do so (Birley & Westhead, 1993, p. 45; Donckels et al., 1987, p. 55). Further, it can be assumed that habitual entrepreneurs, through their previous business ownership experience, (i) have generated wealth that they can use to finance the subsequent venture (e.g., through an exit), or (ii) have a richer social network, and therefore better access to financial funds than novice entrepreneurs (Westhead & Wright, 1998a, pp.

[4]Habitual entrepreneurs agree more with the statements "One of my greatest strengths is identifying goods and services that people want", "I accurately identify unmet customer needs", "One of my greatest strengths is my ability to seize high-value business opportunities", "new business opportunities often arise in connection with a solution to a specific problem", "I enjoy thinking about and / or looking for new business opportunities", "the consideration of one opportunity often leads to other opportunities" (Westhead, Ucbasaran, Wright, & Binks, 2005, p. 121).

179-182). However, findings on using personal funds do not yield a definite result. While Birley and Westhead (1993, p. 45) find habitual entrepreneurs to use family friends and personal funds more, supporting the hypothesis that habitual founders use money from their prior ventures, findings from Donckels et al. (1987, p. 55) indicates the opposite. Habitual also take money from partners (Donckels et al., 1987, p. 55) as well as suppliers and customers (Birley & Westhead, 1993, p. 46) more often, which makes sense considering that novice founders do not e.g., have a network of suppliers at the time of the founding. In terms of subsequent funding from venture capital firms, research shows that habitual entrepreneurs with prior venture capital-backed start-up experience tend to receive higher funding and wait less time until the initial investment (e.g., Zhang, 2011, pp. 17-18). This, in turn, is attributed to differences in human and social capital (Zhang, 2011, pp. 22-23).

2.2.4. Shortcoming of current literature

The previous chapters identify the factors influencing entrepreneurial behaviour and compares novice and habitual entrepreneurs along these variables. However, one should bear in mind that the underlying studies are limited in terms of location and industry sector (see Chapter 2.2). For example, only a few of the findings on personal characteristics from Kolvereid and Bullvag (1993, p. 282) were consistent across different countries. The findings of Paul Westhead and co-authors, who are among the most active researchers in the field studying character traits and behavioural differences between different types of entrepreneurs, are based solely on entrepreneurs from Great Britain and sometimes Scotland only. The extent to which the results can be generalised cannot be assessed due to the lack of comparable studies from other countries.

Besides that, Shane and Venkataraman (2000, p. 218) define entrepreneurship as the interrelationship of "the exploration of the sources of opportunities; the processes of discovering, evaluating and exploiting opportunities; and the group of individuals who discover, evaluate and exploit them", thus highlighting a multifaceted conceptual framework. Moreover, entrepreneurship is emphasised as rather a process than a static state (MacMillan, 1986; Shane & Venkataraman, 2000, pp. 218-219). The classification of entrepreneurs based on prior business ownership experience, however, contradicts these characteristics. First, because focusing exclusively on the entrepreneur as distinguishing variable neglects the multi-layered nature of entrepreneurship (Shane & Venkataraman, 2000, p. 218) and second because classifications require homogeneous, overlap-free groups (Merz, 2008, p. 10), which is not possible considering a process characteristic. For example, considering that all novice founders include aspiring future habitual founders (Westhead & Wright, 1998a, p. 176), researched novice entrepreneurs may have stated habitual characteristics, thus leading to a bias in existing results. Hence, if new entrepreneurs also represent the preliminary stage of habitual entrepreneurs, they represent an unsuitable comparison group (Merz, 2008, p. 12). Consequently, overall, a business ownership-based classification is only of limited use for researching entrepreneurial characteristic differences and the corresponding influencing factors.

Lastly, studies researching differences between types of entrepreneurs focus on founding experience rather than on the necessary resources themselves. In fact: They use stock of business ownership experience as a proxy for entrepreneurial knowledge and skills (e.g., Politis, 2008, pp. 476-478; Westhead, Ucbasaran, & Wright, 2005a, p. 397; Zhang, 2011, p. 9). However, human capital (e.g., Unger et al., 2011, pp. 26-27) and resource-based theory (e.g., Wiklund & Shepherd, 2003, pp. 1308-1310) literature argue that entrepreneurs should have specific knowledge about the industry and business type as well as task-related human capital, that is related to important daily processes (e.g., environmental scanning, opportunity identification, strategy formulation as well as organization, management, and leadership) (Chandler & Jansen, 1992, p. 225; Shane & Venkataraman, 2000, pp. 220-224). The current classification of novice and habitual entrepreneurs based on prior business ownership experience, however, fails to distinguish between prior business ownership experience and other types of experience in which the same skills and knowledge (e.g., opportunity identification or managerial skills) and thus relevant knowledge resources may have been developed.

2.3. Investor-entrepreneur

Private equity refers to an asset class that is not publicly traded on a market (Cumming, 2012, p. 1). Thereby, in general linguistic usage and scientific literature, a distinction is made between venture capital and private equity based on the age of the respective company funded (e.g., Harris, Jenkinson, & Kaplan, 2014, p. 1; Metrick & Yasuda, 2010, pp. 2303-2304; Wright & Robbie, 1998, p. 522). Venture capital thereby includes capital for early-stage and growth projects, whereas private equity mainly includes later-stage growth capital, various types of mezzanine, buy-out investment and turnaround capital (Berk & DeMarzo, 2016, pp. 862-865; Cumming, 2012, p. 1; Wright & Robbie, 1998, pp. 521-522). In this paper, investor defines a decision taker in a venture capital or private equity firm who allocates money to specific investment opportunities.

The private equity investment process typically comprises the steps deal generation, initial screening, valuation, deal approval, execution, and exit (Fried & Hisrich, 1994, pp. 31-32; J. Hall & Hofer, 1993, pp. 27-29). Thereby, the sourcing of potential investment opportunities is an essential step in the investment process (Berk & DeMarzo, 2016, p. 536; Wright & Robbie, 1998), as only a fraction of opportunities meet the investment criteria, represent a potential investment and are therefore considered more closely (Teten & Farmer, 2010, pp. 34-35). In the subsequent valuation of opportunities, different valuation techniques are required for the individual projects, depending on the age of the company to be invested in (Wright & Robbie, 1998, pp. 526-527). Corporate finance literature, therefore, provides a variety

of methods for valuing later-stage companies. Best known are the discounted cash flow (DCF) method, the asset value method and the multiple method (Berk & DeMarzo, 2016, pp. 322-338; Wright & Robbie, 1998, p. 539). However, historical sales figures are often unavailable for early-stage ventures, and statements on cash flows are characterized by uncertainty and potentially rapid growth (Wright & Robbie, 1998, p. 526). As a result, entrepreneurs have information about themselves and their company that is not available to potential investors, creating an information asymmetry between investor and entrepreneur (Shane & Cable, 2002, p. 365; Wright & Robbie, 1998, pp. 524-525). VC investors, therefore, often have to rely on evaluation variables of the founding team such as managerial capabilities, general traits or track record (MacMillan, Siegel, & Narasimha, 1985, pp. 121-123).

The social network is an essential variable for VC or PE investor performance. First, social ties largely contribute to the sourcing of possible deals. In a study analysing eighteen individual investment processes, Fried and Hisrich (1994, p. 31) showed that none of the deals examined were sourced "cold", but all had been entered into through recommendations. Teten and Farmer (2010, pp. 34-35) attribute 45% of deal origination to personal and professional contacts. Following, investors with an excellent social network perform better than those with a comparatively weak one (Hochberg, Ljungqvist, & Lu, 2007, p. 284), highlighting the importance of a good network for the day-to-day business and performance of PE firms. Second, information exchange between social ties is used to overcome the information asymmetries and thus influences the investment decision (Venkataraman, 1997, p. 127). Especially when no public data about the company or entrepreneur is available, investors turn to their social network to get information (Shane & Cable, 2002, p. 366).

With respect to entrepreneurship and considering the above-described characteristics of private equity investing, founders who have previously worked as investors in private equity may have acquired (i) substantial task-related human capital and (ii) the necessary social capital that is important for the foundation and operation of a venture. For example, Wright and Robbie (1998, p. 552) argue that VC investors can have deeper conversations with experienced founders from the beginning on than with their inexperienced novice counterparts. Vice versa, inexperienced entrepreneurs who have previously worked in VC or PE and thus already know the investment process should have, at least in terms of obtaining financial capital, a knowledge advantage over entrepreneurs with other backgrounds. In addition, investors are presented with a multitude of potential investment opportunities on a daily basis and have to assess the feasibility and novelty, growth and potential exit as well as the capabilities of the management team based on the available information (Wright & Robbie, 1998, p. 539). This might positively influence the individual's perception of opportunities and thus the identification of their own business idea. Lastly, because of the high deal flow, PE investors also have above-average contact with founders and other investors with whom they may co-invest (e.g., club deals), which could lead to helpful learnings as well as direct and indirect connections and a resulting rich social network which might help *investor-entrepreneurs* to establish their own venture.

3. Research Method

The entrepreneurial literature has long examined different types of entrepreneurs. So far, differences in decisions and actions between novice and habitual founders are ascribed to differences in human and social capital. However, although the influence and thus the importance of human capital on entrepreneurial processes has been extensively studied and the consideration of task-related human capital is suggested (e.g., Unger et al., 2011, pp. 26-27; West & Noel, 2009, p. 17; Westhead, Ucbasaran, & Wright, 2005a, p. 414), to date only stock of prior (business ownership) experience is considered when comparing different types of entrepreneurs. Different possibilities to acquire the necessary task-related skills and knowledge are entirely neglected. In fact, a distinction between novice and habitual founders based on previous business ownership experience has been established. But what about first-time entrepreneurs who may have acquired the necessary skills, knowledge, and social capital in a profession other than business ownership? Venture capital and private equity investors could be an example. By working a lot with entrepreneurs and experiencing many of the day-2-day tasks from the other side of the table, entrepreneurs might have acquired the necessary human and social capital for a venture foundation. Given the limitations regarding the differentiations along entrepreneurial types and behavioural influencing variables in previous literature, as well as the promising characteristics of private equity investors, I employed an explorative (qualitative), cross-sectional research design to examine my research question: *How do investor-entrepreneurs differ from novice and habitual entrepreneurs?* Thereby, a cross-sectional design represents the data collection from a subset of a population at a single point in time (Bell, Bryman, & Harley, 2018, pp. 53-54). As such, this approach is a commonly used approach in social science research (Bryman, 2006, p. 104) and is particularly applicable for examining gaps in existing literature and identifying new patterns (Bell et al., 2018, p. 53; Edmondson & McManus, 2007, pp. 1160-1162). In doing so, multiple cases are used to achieve variation, examine relationships between variables, and derive general findings with little focus on the unique context of each case (Bell et al., 2018, pp. 53, 63).

3.1. Sample selection

My study included founders who worked in private equity or venture capital before their own founding. Thereby, I only considered private companies that offer innovative goods and services (Venkataraman, 1997, p. 120). I did not include entrepreneurs who subsequently set up their own investment fund or financial advisory company, as they do not

represent a new product or service but rather a new source of existing business.

To identify particular founders, I used Pitchbook, which is one of the most comprehensive and accurate platforms for private companies (Retterath & Braun, 2020, p. 39) and is frequently used for research in the field of VC and PE investments (e.g., Block, Fisch, Vismara, & Andres, 2019, p. 331). In comparison to alternative data sources, the advantage of Pitchbook is that it provides sufficient information about the respective founders in addition to information about the companies (Retterath & Braun, 2020, p. 23). To double-check and expand the information gained from Pitchbook, I also used Crunchbase, which is another comprehensive online database. I first filtered Pitchbook under the category "People" according to the necessary criteria[5]. This search resulted in 760 hits. Second, I only considered products and services that are not in the investment sector (e.g., investment funds, financial advisory firms, money lenders), resulting in a long list of 64 founders. Finally, I conducted the same procedure at Crunchbase[6] and compared the results with the long list from Pitchbook. Through the comparison, I identified another 16 founders.

Since nothing is known about investor-entrepreneurs yet, extensive, detailed, and meaningful data was needed to shed light on the phenomenon (Edmondson & McManus, 2007, p. 1162). Therefore, to identify a manageable number of cases while having sufficiently rich information to investigate key variables and detect patterns, I took a criterion sample (Miles & Huberman, 1994, p. 28) of 12 founders from the population; a number which is recommended for inductive theory building (Eisenhardt, 1989, S.545) and which was feasible considering time and effort. To draw the sample from my population gathered through Pitchbook and Crunchbase, I prioritised the respective founders into three prioritisation groups (A, B, C) according to their time as an investor and the order of decisive career milestones[7]. Even though my focus was on novice founders, I did not exclude habitual investor-entrepreneurs, as the focus of this study was on their first foundation. The founders were contacted successively via LinkedIn and email, starting with group A, until the threshold of 12 was reached based on their feedback.

3.2. Data collection

Consistent with other studies applying an explorative (qualitative) cross-sectional approach (e.g., Charles & Gherman, 2013; Chong, 2008; Durst & Wilhelm, 2012) and recommendations for qualitative research (Bell et al., 2018, p. 57; Edmondson & McManus, 2007, p. 1160), I conducted semi-structured interviews. This type of data collection allowed for flexibility and a smooth, intuitive flow of questions during the interviews (Bell et al., 2018, p. 467). In advance to the respective interviews, I developed an interview guide (Bell et al., 2018, p. 473), including important aspects identified in the theory section (e.g., human capital, social capital, entrepreneurial behaviour). Together with some background information about the study, the guide was sent to the participants before the interview (see Appendix B). All interviews were conducted within a six-week period between May and July 2021. Interviews usually lasted 30-50 minutes with exceptions upwards (60 minutes) and downwards (15 minutes) and were conducted by telephone or video conference. The availability of the interview partners primarily determined the length of each interview. All interviews were recorded, subsequently transcribed, and used as the primary source of information for this study. Thereby, interviews were conducted and transcribed either in German or English, the native language of the respective participants.

To mitigate biases from imperfection recall and to triangulate the information from the interviews (Yin, 2009, p. 102), I collected additional founder-related information about e.g., the age or the previous employers, from the internet, company websites, and LinkedIn.

3.3. Data analysis

To analyse the data obtained through the interviews, I followed a *quasi-inductive*, mixed-method coding approach (Perry & Jensen, 2001, p. 4). This type of coding is a modified form of inductive coding (Corbin & Strauss, 2008, p. 104) often applied in the context of explorative research (e.g., Preller, Patzelt, & Breugst, 2020, p. 5). It allowed me to establish pre-categories in advance based on existing literature, which helped me to first become aware of several important dimensions of the phenomenon to be researched. However, the pre-categories were not meant to test the existing theory, but to "re-test" it in a real-empirical setting, to re-specify, refine, or eliminate it contextually. This allowed me to link the specific principles of grounded theory directly to the existing literature (Perry & Jensen, 2001, p. 4).

In line with this approach, I fist created an initial coding scheme (Bell et al., 2018, p. 299) based on findings from previous studies at the beginning of the coding. The original scheme consisted of three main distinguishing elements, eight first- and eight second-order codes that I thought were important to distinguish the entrepreneurial types (see Figure 2). In the actual analysis process, I first compiled general information on each case in a tabular overview (Miles & Huberman, 1994, p. 90). Next, I applied open coding (Corbin & Strauss, 2008, pp. 222-223) and assigned data segments to existing and emerging categories (e.g., fundraising, social capital, opportunity identification), using the MaxQDA software to manage my data. Although I had pre-determined codes, I approached the individual cases with an open mind (without any foreseen propositions already in place) and thus allowed the data to speak for themselves (Suddaby, 2006,

[5]Pitchbook criteria: company type (investors: PE/Buyout & Venture Capital) position status (active & former) main position (founders & co-founders) and location (Germany).

[6]Crunchbase criteria: location (Germany), primary job title (founder), past job (includes any: venture capital)

[7]E.g., assuming that the influence of learnings from a previous job is greatest when directly applied in the next job, founding directly after the job as an investor was seen as more valuable (Reuber and Fischer (1999, p. 39) argue that knowledge only has a certain shelf time), while stints apart were seen as less valuable; a minimum of two years as an investor was required to assume a certain amount of knowledge and skills acquired.

p. 634). In this way I was able to discover new codes and the coding scheme was continually adapted. In doing so, I constantly compared the codes to all the cases and reviewed them to see which codes matched multiple cases and which were unique (Glaser & Strauss, 2017, pp. 105-113). Subsequently, I applied axial coding (Corbin & Strauss, 2008, p. 165) where I compressed the categories and started to seek suitable literature for the preparation of my data and to compare initial results with theory. Thereby I also considered constructs from existing literature to condense codes (e.g., task-related HC). As a final step, I assigned the resulting codes to the main distinguishing elements and dropped codes for which I did not have efficient enough information, thus applying selective coding (Vollstedt & Rezat, 2019, p. 89).

Lastly, I created quotation tables for each of the relevant first and second order codes, consisting of short blocks of text with respective relevant statements from the interviewees (Miles & Huberman, 1994, p. 93). These were primarily intended to help structure the answers of the individual founders and to conduct cross-case comparisons (Eisenhardt, 1989, p. 540; Miles & Huberman, 1994, pp. 172-177). My aim was to identify possible relationships between the distinguishing characteristics (Eisenhardt, 1989, p. 541) and to derive clusters or groups that have these characteristics in common (Miles & Huberman, 1994, p. 174). In doing so, I observed multiple similarities and differences between cases. I grouped the similarities into categories and compared them with the existing literature. Focusing on differences, I examined the respective cases in more detail and found a link between deviating cases and specific case dimensions. As a result of comparing the respective results with existing research and compare emerging theory with all cases in multiple iterations, I developed propositions with regard to differences of the investor-entrepreneur and the respective variables influencing them (Miles & Huberman, 1994, pp. 75-76).

4. Results and discussion

To reach the sample of twelve founders, I had to contact a total of 52 founders. Reasons for this were mainly refusals and non-responses of prioritised founders. The most common reasons for rejection were lack of time, general unwillingness to participate in academic studies and no follow-up response after initial commitment. Finally, one interview turned out to be not very rich in information. Therefore, I conducted an additional interview (in sum 13).

My final coding scheme comprises eight first- and fourteen second-order codes and only includes variables that I consider potentially relevant to answer the research question of this study (see Figure 2). In the course of coding, I dropped three first-order codes and two second-order codes (dashed lines) because I was not able to gather enough information from the interviews to gain meaningful insights into the respective topic. The corresponding insights from the interviews are explained below, summarised under the individual main distinguishing elements.

4.1. Findings with regards to human capital

> *"I think that was a warm start compared to a founder who has never had anything to do with VCs." – David, V3*

Since human capital is supposed to have significant effects on the performance of a founder and thus his company (Unger et al., 2011, p. 23), I was mainly interested in the founder's initial resources and stock of experience as well as on specific HC while analysing possible differences between investor-entrepreneurs and other types of entrepreneurs. Thereby I was especially interested in possible task-related HC necessary for founding a venture. This leads to the following findings.

First, based on descriptive statistics (see Table 1), insights about the initial resources (age, gender, education) of the founders in this study were obtained. Thereby, the average age at the time of founding is 30 across the sample, while investor-entrepreneurs with a VC background (avg. 28) appear to be significantly younger compared to their PE counterparts (avg. 32). Compared to earlier findings from Kolvereid and Bullvag (1993, p. 279) or Birley and Westhead (1993, p. 43), the investor-entrepreneur characteristic do not match with other types of entrepreneurs. Overall, investor-entrepreneurs tend to be significantly younger than "normal" novice ones. In terms of gender, a clear dominance of male investor-entrepreneurs can be observed (Table 1). Twelve out of the thirteen founders interviewed were men (92%), showing similarities to findings from earlier studies (Birley & Westhead, 1993, p. 44; Kolvereid & Bullvag, 1993, pp. 278-279). Even beyond the sample, the number of female founders with corresponding criteria was deficient. Of the total 52 founders contacted, only three were female. A possible explanation for this is the generally low proportion of women in the venture capital and private equity industry (Cassala, 2018). Thus, the basic population of potential female investor-entrepreneurs is inherently small. Examining the educational background shows that eleven out of thirteen founders, regardless of their background, have a master's degree (or comparable e.g., diploma). Even though previous studies are not consistent in finding significant differences between novice and habitual entrepreneurs, the data from this study suggests that investor-entrepreneurs have a higher level of education than other novice and habitual counterparts (cf., Birley & Westhead, 1993, p. 44; Donckels et al., 1987, p. 52; Kolvereid & Bullvag, 1993, p. 279).

Second, regardless of whether the founders in this study have VC or PE experience, on average, they work for three organisations before founding a company themselves (Table 1), thus showing the same stock of experience as novice founders in other studies (cf., Birley & Westhead, 1993, p. 43; Westhead, Ucbasaran, Wright, & Binks, 2005, p. 112). On the other hand, all founders, except of Nate (P4), were employed in companies with <1,000 employees. For founders with a VC background, it is even mostly companies with less than 100 employees; thus, with regards to that, they tend to show habitual characteristics (cf., Birley & Westhead, 1993, p. 45).

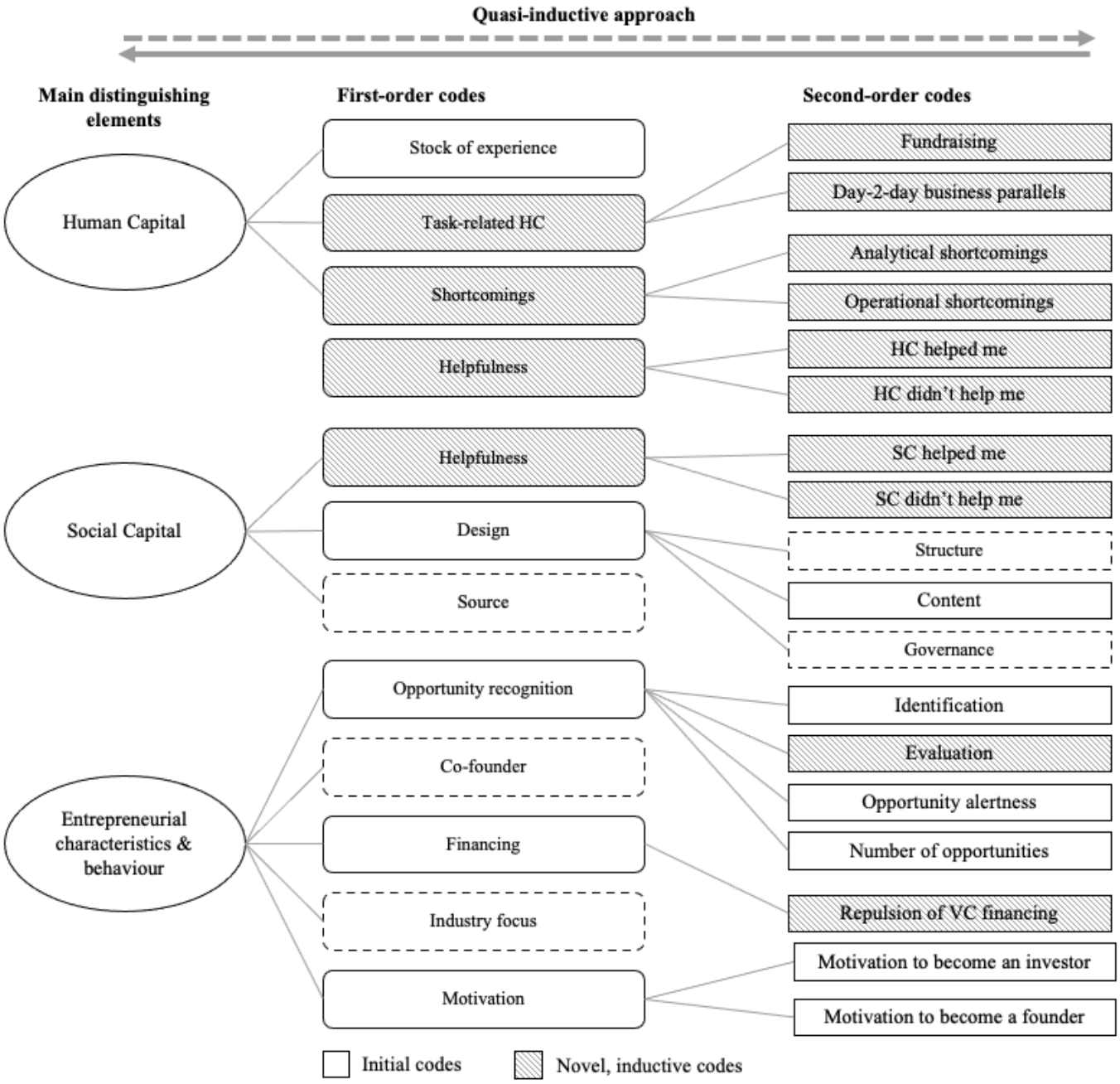

Figure 2: Coding scheme.

Third, regarding acquired skills and knowledge (HC investment outcomes), the results of this study suggest two main findings: On the one hand, acquired skills between VC and PE investors differ in terms of their type and usefulness for a subsequent founding process. On the other hand, previous experience as an investor has a generally positive effect on the skills and knowledge required for the acquisition of external capital. Table 3 (Appendix C) summarises the relevant passages concerning key skills and knowledge acquired as an investor. While VC founders emphasise "define business model", "learn dos and don'ts", and "general early-stage venture understanding", PE founders tend to mention "financial modelling", "entrepreneurial understanding", and "opportunity evaluation". Interestingly, the statements reflect a particular phase commitment of the individual founders. The statements from VC founders clearly indicate helpful HC in the early phase of venture creation (e.g., business definition and general process understanding at the beginning), while PE founders emphasise necessary HC for later stages of a company (e.g., financial modelling or entrepreneurial understanding). These statements align with the previous job focus, i.e., VC equals early-stage investments, PE equals growth investments. However, the respective founders see these acquired skills as differently helpful for founding a ven-

Table 1: Overview of interview partners.

Founder		Age	Gender	Entrepreneurial type	Education		Prior job				Industry focus of first venture	Year founded first business	Age at first foundation	Stock of experience*
					Highest degree	background	type	number of years	number of employees	Industry focus				
Anton	V1	31	male	portfolio	M.Sc.	Management	VC	3,5	5-10	Fintech, Insurtech, Regtech, Mobility	Fintech	2019	29	1
Charly	V2	32	male	serial	B.Sc.	Corporate Finance, Computer Science	VC	2	5-10	Medtech, Industry 4.0, IoT	Medtech	2016	27	5
David	V3	32	male	novice	M.Sc.	Mechanical engineering & Business Administration	VC	3,5	10-15	Medtech, Biotech, Pharma, Software, Mobility	Legaltech	2018	29	2
Justus	V4	27	male	novice	M.Sc.	Computer Science	VC	2,75	10-15	Tech, Cloud, Semiconductor	Development automation	2020	26	4
Leo	V5	29	male	novice	M.Sc.	International management & Financial management	VC	2,5	40-50	Industry 4.0	Proptech	2020	28	3
Mike	V6	37	male	portfolio	M.Sc.	Finance & Marketing	VC	4,5	100-150	multiple	Wholesale	2013	29	2
Robby	V7	28	male	serial	M.Sc.	Business Administration, Analytics & Big Data	VC	2	100-150	multiple	Gambling	2020	27	4
Steven	V8	31	male	serial	M.Sc.	Business Administration	VC	2	25-30	Fintech & B2B Software	Building management	2021	31	2
Felix	P1	36	male	novice	M.Sc.	Mechanical engineering	PE	6	20-30	Consumer, Technology & Software, Service	E-Commerce	2020	35	2
Ida	P2	29	female	novice	B.Sc.	Mechanical engineering & Finance	PE	2	400-500	multiple	Procurement software	2020	28	1
Max	P3	39	male	serial	M.Sc.	Business Administration	PE	2,5	230-240	Small & Mid Cap (all industires)	Marketing automation	2014	32	2
Nate	P4	31	male	novice	M.Sc.	Mathematics & Finance	PE	5	1,200-1,300	multiple	Fintech	2018	28	4
Peter	P5	41	male	novice	M.Sc.	Management	PE	7	130-140	Industrials, Consumers, Services	Financial services	2019	39	4
Average		33						3					30	3
Average VC		31						3					28	3
Average PE		35						5					32	3

*Number of companies before founding venture; inchuding PE/VC companies

ture. While David (V3) considers it generally helpful ("I think that was a good basis for taking on this role at [...]"), Anton (V1) limits it more to the non-operational activities ("If I negotiate a participation agreement with the investors now, then I have done that before. But the operational things; those are rather necessary administrative things to finance the operational business."). Peter (P5), however, does not consider it helpful at all: "So for the foundation it didn't help me." Overall, six interview partners (V2, V3, V4, V5, V6, P1) stated that it was very helpful (high), three (V7, V8, P2) that it had an impact (medium) and three (V1, P4, P5) that it was not helpful (see Table 5, Appendix C). Interestingly, only one of the interviewees with a PE background stated that it was very helpful, while five out of eight founders with VC background did so.

Further, the results of this study suggest that prior work experience as an investor can positively affect the skills and knowledge related to the acquisition of external capital. All interview partners reported relevant learnings they had acquired as investors, summarised under "general process understanding" and "key evaluation criteria". Table 4 (Appendix C) shows the respective statements. "General process understanding" thereby covers the process of a private equity or venture capital investment process, whereas "key evaluation criteria" describe the points of reference in the evaluation step of the funds. For example, David (V3) shares his learnings regarding what early-stage investors look for: "[. . .] i.e., what kind of people do you have to have, what is a good team, what is a good sales development, what is a good vision, what is an exciting technology" and that this knowledge helped him in with his own venture: "That helped me especially when I did fundraising". He further backs this up by describing the fundraising process as a "warm start compared to a founder who has never dealt with VCs before", while Leo (V5) describes the process without prior experience as a "black box". Overall, while previous studies only consider prior venture capital-backed business ownership experience as an influencing variable for venture capital acquisition (e.g., Zhang, 2011, p. 15), the results of this study suggest that work experience as an investor can have similar effects on necessary skills and knowledge (task-related human capital).

Fourth, although the founders in this study emphasize that previous experience in VC or PE does imply helpful HC (e.g., fundraising), they also point out possible shortcomings that go hand in hand with the job as an investor. Justus (V4) and Paul (P4), for example, mention a lack of people management ("The steps are less clear and you have to learn how to develop people, how to manage people and that is not taught in PE as well as VC." – Justus, V4; "Where you definitely have shortcomings is the topic of team management, that whole management topic." – Peter, P5), while Mike (V6), Robby (V7) and Steven (V8) mention more general shortcomings in operational skills ("I think you sometimes underestimate the operational complexity of setting up a company." – Mike, V6; "So I think [. . .] that if you come from VC, you might not be as good operationally" – Robby, V7; "I think what is underestimated a bit is the switch from investor to operator." – Steven, V8). The founders cite different reasons for this (see Table 2). Thereby, since it was mentioned twice, "high performance environment" refers to the working culture of the previous job as an investor. Here, the main focus of respective funds is on independent working styles and individual independence, which does not necessarily require structured personnel management (e.g., "[You're] working with "high potentials" all the time and sometimes it doesn't matter how they're handled" - Peter, P5). Thus, the respective potential founders learn nothing in the area of person-

Table 2: Reasons for entrepreneurial shortcomings.

Investor	Reason for shortcomings	Statements regarding the reason why investor entrepreneurs have shortcomings
V5	High performance environment	In PE or consulting, you're only surrounded by really great high performers all the time. Hopefully in your start-up too, but that's just something different.
V7	Lack of thematic deep dives	[...] you might not be so good operationally, because you're looking at things at a very "high level", which is often just a herd game. They are always chasing the latest trends and it is very rare that you do a real "deep dive".
V8	Lack of value creation	As an investor, you don't actually produce any output and you don't actually do anything, if we're being honest. You're like an information machine, you kind of sit there, listen to pitches all day and then make a lot of decisions about what you're going to do. [...]. And that's just something completely different as an operator. It's actually about doing things and doing things somehow quickly and testing things and somehow producing output.
P5	High performance environment	Because in PE or investment banking, you're always dealing with high potentials, and sometimes dealing with them doesn't matter, but when you have a company with 20 or 30 people, you don't just have these extremely high potentials, but also normal people, and you need structures and weekly meetings; one-to-one meetings, and dealing with these profiles is important, and you simply don't learn that in PE.

nel management, which ultimately leads to shortcomings in their own later founding. Second, "lack of thematic depth" and "lack of value creation" can be summarised in terms of content: The job as an investor has no relation to operational tasks. This consequently leads to operational shortcomings. Third, Mike (V6) and Steven (V8) point out that former professional investors sometimes approach entrepreneurial tasks too analytically: "I think the flip side of it is that you almost think too much because as an investor, especially as a VC, your job is to find reasons why something might not work and that's just the opposite of the entrepreneurial mindset. Whereas an entrepreneur often just; yes, often it's just people do it and change the business model along the way or it just doesn't work or they're lucky and it works very well. And you just don't have that." – Steven, V8; „I would say in the negative that they then must learn that not everything can be done by analysis, but that sometimes you have to force your luck. There is often too much emphasis on analysis and not so much on doing." – Mike, V6. Thus, according to Mike (V6), the job as an investor might cause a "paralysis through analysis". These statements indicate a contradiction with the focus on execution that is considered necessary in the entrepreneurial mindset literature (e.g., McGrath & MacMillan, 2000, pp. 3-6).

Considering the insights regarding human capital and the respective characteristics, I propose:

Proposition 1a: Investor-entrepreneurs acquire business ownership relevant skills and knowledge (e.g., fundraising, or helpful dos and don'ts) through their job as investor, thus mitigating a lack of business ownership experience.

Proposition 1b: Only human capital acquired as a venture capital investor is helpful for a subsequent venture foundation.

Proposition 1c: Prior experience as an investor causes shortcomings in the areas of people management as well as in the entrepreneurial mindset.

4.2. Findings with regard to social capital

"So, what definitely helps a lot is the network you build." – Justus, V4

In examining the individual founder's social capital, I was interested in the personal network of the individual founders and its usefulness. The aim was to determine whether the previous job as an investor impacted the subsequent venture and the founding process more precisely. Following Mosey and Wright (2007), I was mainly interested in the network's structure, the content acquired through the network and the individual network maintenance. However, not all aspects yielded enough information or significant insights. Thus, only the network content (cf., Mosey & Wright, 2007, p. 912) could be evaluated. In addition, helpfulness proved to be an often-mentioned code. The following three points summarise the resulting insights.

First, individual statements regarding the helpfulness of the network acquired as an investor for founding one's venture reveal significant differences between VC and PE founders. Table 5 (Appendix C) summarises the corresponding statements. While eight of the founders describe it as very helpful ("high"), four founders say it is not helpful ("low"). Thus, no significant tendency can be observed, and no general statement for all investor-entrepreneurs can be made. However, looking at the individual sub-groups, six of the eight VC founders (V2, V4, V5, V6, V7 and V8) consider their acquired network very helpful, while only one out of five founders with a PE background (P5) consider it helpful. In conclusion, VC experience contributes to a more helpful network development for a subsequent venture foundation. This is further illustrated by founders with PE backgrounds who directly address the difference. For example, Max (P3)

says: “I say VC would have been more relevant than PE” or Felix (P1), who refers to his co-founder: „I notice this a lot; [Name of co-founder] worked in VC all the time, and his network is of course super helpful for everything we do here.“ Again one possible explanation could be the stage focus of the individual founders in their previous job. While VC founders deal a lot with early-stage companies and thus also with possible contacts in this area, PE founders tend to build up a network with contacts relevant for later stages of a company (e.g., Felix (P1): "You have a relatively strong network in the direction of M&A advisors, lawyers, banks and so on").

Second, investor-entrepreneurs show similarities with "normal" novice founders in terms of network content, although the focus is somewhat different. To discover possible differences, I summarised and analysed the individual statements of the founders and then clustered them (see Table 6 – Appendix C). In contrast to the different statements of the founders regarding the influence of social capital on the founding of a venture, statements regarding the content acquired through the network were more aligned. Thereby, obtaining “feedback on the initial idea during the beginning of the entrepreneurial process” is the most frequently mentioned reason (six founders), followed by “hiring” (five founders) and “personal introductions to third parties” (four founders). “Funding”, “market knowledge”, and “customer acquisition” were mentioned twice by the founders, while “co-founder” and “further opportunities” represent unique cases. Hence, if investor-entrepreneurs must be classified based on previous studies, these entrepreneurs show similarities to their "normal" novice counterparts who mainly use their network to engage in the next stage of business development, to seek proof-of-concept financing, industry knowledge and potential customers, even if these use cases correspond to the rather less frequently mentioned statements of the investor-entrepreneurs (cf., Mosey & Wright, 2007, p. 927). Further, "personal introduction to third parties" shows parallels to other novice academic entrepreneurs. For example, Mosey and Wright (2007, p. 928) show, that novice entrepreneurs use the university’s TTO (technology transfer officer) as a broker to build new relationships, thus using a third party to build new ties. However, one should bear in mind that the study by Mosey and Wright (2007) only considers academic entrepreneurs. Even though “feedback on the initial idea during the beginning of the venture process” is not mentioned in their study and thus a direct comparison is not possible for this point, it can be assumed that this represents a typical process that founders with other backgrounds and thus "normal" novice entrepreneurs also do. For example, Leo (V5) comments as follows: “I think everyone uses their network, so if you’re not starting up straight out of university, then no matter what you’ve done before, you’re going to activate your network as much as you can.” Accordingly, even if the acquisition of initial feedback can be observed among other types of entrepreneurs, it represents a unique feature of investor-entrepreneurs until proven otherwise. Overall, however, investor-entrepreneurs show similar characteristics to their “normal” novice counterparts.

Third, although several founders report shortcomings in the area of team leadership and operational skills in Chapter 4.1, only Leo (V5) states that he has obtained missing human capital through his network („[. . .] that also helped me enormously to exchange ideas with other founders [...] and they opened many doors for me to others, and I have to say that I had almost no real estate experience before, so I first had to build up a bit of experience myself“). Indeed, investor-entrepreneurs seem to use their network less to replace missing human capital (cf., West & Noel, 2009, p. 9). One reason for this may be that all founders have co-founders, thus they may have already replaced missing HC with their co-founders before launching. Although this does not indicate “normal” novice characteristics, it contradicts typical habitual ones (i.e., habitual entrepreneurs specifically use their network to fill resource gaps – cf. Mosey and Wright (2007, p. 929)) and therefore highlights that investor-entrepreneurs cannot be clearly assigned to existing types of entrepreneurs.

Considering the insights regarding human capital and the respective characteristics, I propose:

Proposition 2a: Venture capital experience contributes more positively to the development of social ties necessary for a subsequent foundation of a venture than PE experience.

4.3. Findings on entrepreneurial characteristics and behaviour

> *“Well, I mean, I think it’s clear why you found a company: First, because it’s fun. Second, because I think it’s cool to build something yourself, to build a team, to really be responsible for something, and third, there’s clearly the monetary aspect.” – Robby, V7*

In examining the entrepreneurial characteristics of investor-entrepreneurs, I followed prior studies (e.g., Birley & Westhead, 1993; Donckels et al., 1987) on differences between novice and habitual entrepreneurs and focused on the topics of opportunity identification, opportunity alertness, motivation and financing. Nevertheless, during the course of coding also new sub-topics emerged which are considered in the respective sub-chapters. Cognition as a heavily studied topic in the field of entrepreneurial behaviour (e.g., Baron, 2006; Reuber & Fischer, 1999; Westhead, Ucbasaran, & Wright, 2005b) was not explored due to the scope of this thesis and the difficulty of measuring it in a qualitative study. Overall, the findings of this study indicate that investor-entrepreneurs differ from "normal" novice entrepreneurs in terms of entrepreneurial traits. The corresponding characteristics and behaviours are discussed in the following section.

4.3.1. Opportunity identification, evaluation & alertness

By analysing how the individual founders identified their idea, I quickly noticed that they mentioned typical characteristics in terms of opportunity identification and “investor specific” ones. Moreover, when asked about the origin of their business idea, the respective founders were happy to

elaborate a little further. Therefore, I summarised the corresponding statements, analysed them and condensed them to essential key take-aways (see Table 7 – Appendix C). From the respective statements, I was able to derive insights about the sources of information used for identification and was accordingly able to compare the founders with characteristics of other types of entrepreneurs. Hence, four main patterns could be observed: "Topic viewed as an investor", "Pain point in daily business", "Being approached as co-founder", and "actively looking".

First, typical characteristics including "pain point in daily business" and "active looking" reflect "classical" patterns in connection with identification (cf. Westhead, Ucbasaran, Wright, & Binks, 2005, p. 121). Here, the founders state that their idea for the respective venture was based on a problem in their (professional) everyday life or a structured analysis process. David (V3), for example, says "We were very frustrated by it as clients", while Steven (V8) indicates a conscious "top-down" process: "And now this story here with [...] that is "top-down" analysis, so to speak." The corresponding sources of information also match the typical sources used by founders of different entrepreneurial types. David (V3), for example, had a "talk with [his] co-founder", while Felix (P1) says he "chatted with a buddy", thus obtaining necessary information through family and friends. Ida (P2), on the other hand, talked to other business professionals about the perceived problem or opportunity, thus using "customers" and "other business owners", which is also used by other types of entrepreneurs: "Then I talked to a lot of people from the industry - so about 40-50 actually" (cf., Westhead, Ucbasaran, & Wright, 2005a, pp. 403-412).

Second, special characteristics including "topic viewed as an investor" and "being approached as a co-founder during the investor time" represent identification processes that do not fit into patterns of previous studies and partly represent a unique identification process of investors. "Topic viewed as an investor" thereby indicates that the respective founders acquired the idea for their own venture in the course of their work as an investor. While Anton (V1) and Justus (V4) identified the idea in the screening process, Charly (V2) mentions the complete VC process and the sector focus of the fund. Thus, it seems that the job as an investor gives access to superior information regarding technologies and markets. Following research in the area of opportunity identification, the ability to identify entrepreneurial opportunities is influenced by information, know-how and skills that contribute to the processing of such information (Shane, 2003, pp. 45-50; Venkataraman, 1997, p. 124). Interestingly, all founders who expressed this opinion founded their company with the same industry focus as their previous fund (see Table 1). Therefore, we can assume that the respective founders had the necessary information and the corresponding skills, insights, and expertise to identify the corresponding opportunities. Justus (V4), for example, responds to the question where he sees parallels between the job as an investor and entrepreneurship: „That's where you learn to assess market opportunities, rate market opportunities [...]". Thus, the previous employer, namely investment fund, provides the founders with the information, skills, and experience necessary to identify entrepreneurial opportunities. Interestingly, however, it is only founders with VC experience who describe this type of opportunity identification, indicating a VC-specific phenomenon.

Further, "being approached as co-founder" describes the fact that friends, business contacts or even other entrepreneurs approached the respective founders and asked if they would like to implement the other person's idea as a co-founder. While Robby (V7) says very openly and directly: "I didn't have much to do with it. It all came from [name of co-founder]", Max (P3) describes it more impartially "But then [name of co-founder] said 'Hey, I'm going to found this now' and I found it exciting and joined in". Leo (V5), however, got in touch with his co-founders because they were looking for funding from his previous fund. Eventually, they took him on board as co-founder for the new idea in Germany. Thus, this type of identification does not use any specific information but rather refers to opportunities that arise from the job as an investor or the network that comes along with it. This corresponds to the view that socially provided information and thus social capital is indeed helpful in identifying business opportunities (cf., Ozgen & Baron, 2007, p. 186), and indicates that investor-entrepreneurs have superior social capital. However, while five out of seven VC founders mention these methods of opportunity identification, only one with PE background does. This highlights the assumption already made in Chapter 4.2 that only venture capital experience positively impacts social ties necessary for a business foundation.

The suggested superior human (Chapter 4.1) and social capital (Chapter 4.3.1) is also reflected in the perception of opportunities. Regardless of the approaches and information used to identify the respective business idea, most founders claim to have an opportunity alertness. Seven out of nine founders who were asked the question claim to have this characteristic (see Table 8 – Appendix C). Since this quality is influenced by human as well as social capital (cf. Venkataraman, 1997, p. 122) and comparatively "normal" novice founders identify themselves only partially or not at all with an opportunity alertness (cf. Westhead, Ucbasaran, Wright, & Binks, 2005, p.121), this indicates once again superior human as well as the social capital of investor-entrepreneurs. However, a closer look at the sub-groups again reveals a difference between PE and VC investors. While all the interviewed investors with a VC background state that they have an opportunity alertness, only two out of four investors with a PE background claim this (P2, P3). This is in line with earlier findings of this study regarding the helpfulness of social capital acquisition as a VC investor (cf. Chapter 4.2) and supports that previous experience as a VC investor contributes positively to the development of beneficial founder-related characteristics. Moreover, given answers to the question of further identified opportunities further support this assumption (see Table 9 – Appendix C). While four of the seven founders with a VC background

stated that they had identified more than two further opportunities, there were zero among the founders with a PE background. Furthermore, none of the VC founders stated they had not identified any further opportunities, whereas Felix (P1) and Peter (P5) had. Overall, compared to findings from Westhead, Ucbasaran, and Wright (2005a, p. 408), most VC investor-entrepreneurs thus show habitual characteristics in terms of opportunity alertness and number of opportunities identified.

In the course of opportunity identification, the respective founders also repeatedly mentioned the subsequent evaluation process. Thereby I could observe that this process is usually described as very structured and well-thought-out. Table 10 (Appendix C) summarised respective statements. Especially "validation with other market participants", and "precise analysis of the market environment" were referred to several times. However, also "cost-benefit analysis", "validation through known patterns" or "hypothesis driven" describe rather structured than intuitive evaluation processes. Robby (V7) even describes it as a highly structured process with predefined metrics, which he adopted from his time as an investor: "You have various metrics, such as the business model, the fit to the fund, the team, the product-market fit, the timing, and [...] they are always ranked from one to five". Although previous studies researching decisions and action (e.g., Westhead, Ucbasaran, & Wright, 2005a) do not consider the evaluation process and therefore a direct comparison with other types of entrepreneurs is not possible, these findings suggest that former professional investors analyse their own opportunities more genuinely and in a more structured way than their counterparts, reflecting their former evaluation process. This is supported by Steven (V8) who notes that former professional investors are better at "[thinking] in advance about exactly what kind of business model you are doing". Further, these insights reinforce findings from Chapter 4.1 that investor-entrepreneurs somewhat infringe on the entrepreneurial mindset when accessing opportunities.

Considering the insights regarding human capital and the respective characteristics, I propose:

Proposition 3a: Human and more specifically social capital acquired as venture capital investor positively influences the opportunity perception and alertness.

Proposition 3b: Venture capital funds provide necessary information as well as skills and knowledge to discover entrepreneurial opportunities, thus providing potential founders with their subsequent business ideas.

Proposition 3c: Investor-entrepreneurs evaluate their entrepreneurial opportunities using market and information-based valuation variables they acquired through their job as investor.

4.3.2. Motivation

To better understand the behaviour of investor-entrepreneurs and to work out possible differences compared to other types of entrepreneurs, I researched their motivation in a broader sense. Therefore, I was interested in the motivation to become a professional investor first before I also researched the motivation to start one's own business. The broader scope of the study enabled me to conclude the characteristics of the respective founders.

By analysing the motivation to become a professional investor, I was able to identify six core motivations (see Table 11 – Appendix C). Thereby, the individual motivations differ along each other. For example, some founders mentioned the desire to have a more entrepreneurial working style (e.g., Max – P3: "I had the hope that there would be less financial engineering and more really operational work with the portfolio companies."), other emphasized the need to work on more innovative topics (e.g., David – V3: "I wanted to be on the road where innovation really takes off in the market"). For others (e.g., Leo (V5) and Mike (V6)), no specific core motivation could be identified. Overall, no generalisable statements can be made for investor-entrepreneurs regarding their motivation when looking at the complete sample. However, when looking at the respective sub-groups, four individual motivation patterns can be observed for individual investor jobs. First, the reason "work on innovative topics" is stated exclusively by founders who have chosen a job as a VC investor. Here, a possible reason for this could again be the stage focus of the individual funds. While PE mainly deals with established companies, VC covers early-stage ventures and thus novel topics. Therefore, this motivation distribution is not surprising but rather confirms the right job choice of the individual founders. Second, three VC investors (V1, V3, V8) say they chose the job as an investor as an intermediate step on the way to their own venture. Steven (V8) for example says: "I've always wanted to do something on my own and have seen all the steps up to this point only as intermediate steps". Reasons given for this intermediate step included the additional time and the learning curve, which help in "maturing" one's own skillset (Anton, V1) as well as one's own idea (David, V3). Hence, some investor-entrepreneurs choose a job in a VC firm as a precursor to starting their own business, as they do not feel confident enough right after university (e.g., Anton (V1): "I just didn't want to do it straight after my studies. Sure, you could have done that, but your just unexperienced"). Third, almost all PE founders (except Ida, P2) name entrepreneurial working style as a core motivation for becoming an investor. This is interesting because becoming an investor in the first place means deal generation, evaluation and execution (cf. Chapter 2.3) and not operational business tasks. However, founders cite small funds (P3), flat hierarchical structures (P1) and investing their own money (P5) as important factors in the decision to become a PE investor, underlining the founders' desire to be involved in influencing things themselves. Fourth, all PE founders state "university interest in finance" or "learn about finance/investing" as core motivation to become an investor. This is interesting, as only one founder with a VC background states that he chose the job as an investor in order to work in finance and thus learn more in this area. This illustrates a general financial affinity of PE founders and concludes that

PE founders choose the job as an investor because of the opportunity to deepen their financial skills.

In contrast, results from analysing statements regarding the motivation to become a founder reveal that investor-entrepreneurs mostly choose to become a founder for monetary reasons and the desire for self-determination as well as to create values (see Table 12 – Appendix C). Indeed, eight out of ten founders throughout the whole sample cite monetary incentives as a core motivation to become a founder. Although financial motivations have been identified as important in previous studies, the respective statements related more to "financial independence" or "give myself, my spouse, and children security" (Donckels et al., 1987, p. 53; Westhead & Wright, 1998a, p. 189). More specifically, "high income" was rated as not so important in earlier studies (Donckels et al., 1987, p. 53). Therefore, this motivation represents a rather unique characteristic of investor-entrepreneurs. "Self-determination" as the second most frequent motivation is mentioned by both PE and VC founders and can be equated with "need for independence" (Westhead & Wright, 1998a, p. 189) and “personal independence” (Donckels et al., 1987, p. 53) due to adjectives such as "independently, freely, on my own" (David, V3). Therefore, these core motivations do not represent unique characteristics. "Create values" is also mentioned by PE and VC founders, but with a majority of PE founders (three out of four founders with PE background). Besides the fact that “create values” is not included as a criterion in previous studies and hence a comparison with other types of entrepreneurs is not possible, these findings support the insights already gained from the reasons why founders choose the job as a PE investor - namely the urge to be entrepreneurial. Finally, "flattening learning curve" is only mentioned by Anton (V1) and David (V3), thus indicating a typical phenomenon for venture capital investors, and is also reflected in shorter retention periods as investors compared to founders from PE (see Table 1). Interestingly, when asked about the reason for becoming an investor, the respective VC founders (i.e., Anton and David) stated, among other things, "intermediate step towards own founding". Thus, the flattening learning curve can be seen as a starting signal for the active search or special awareness of opportunities and thus as a motivation to become a founder (e.g., Anton (V1): “But at some point, after two years, the learning curve flattened out and then I thought about doing something of my own.”)

Overall, investor-entrepreneurs name core motivations comparable with other types of entrepreneurs (e.g., need for independence), and unique ones (e.g., monetary incentives). Based on these findings I propose:

Proposition 4a: Venture capital experience can help potential founders prepare for the tasks as entrepreneurs and mature possible business ideas.

Proposition 4b: Investor-entrepreneurs have an above-average desire for financial compensation. Thus, financial incentives are one of their core motivations to become a founder.

4.3.3. Financing

While asking the individual founders how they financed their respective businesses in the beginning, business angels, family and friends and own funds turned out to be the most frequently used forms of venture financing (see Table 13 – Appendix C). In total, six different types of financing could be identified, with subsidies (V3), bootstrapping (V4) and family offices (V8) representing unique cases. “Family and friends” as a frequently used form of financing thereby is in line with the sources used by "normal" novice and habitual founders (cf., Birley & Westhead, 1993, p. 45; Donckels et al., 1987, p. 55). “Own funds” also represent a type of financing frequently mentioned by other types of entrepreneurs (cf., Birley & Westhead, 1993, p. 45; Donckels et al., 1987, p. 55). The literature argues that habitual state this more often due to possible financial resources from an earlier exit (cf., Westhead & Wright, 1998a, p. 181). Although this cannot be the case for novice investor-entrepreneurs in this sample, the still very high statement rate suggests that the respective founders earned enough in their previous profession to build up savings that they subsequently used for their own venture. This once more highlights the above-average salaries mentioned in Chapter 4.3.2.

Interestingly, business angels as the most common form of financing are not directly considered in prior studies, making a direct comparison impossible. However, this type of business financing could be included under "other sources" in previous studies (Birley & Westhead, 1993, p. 46; Donckels et al., 1987, p. 55). However, since only 5% of founders rely on such "other sources", business angels do not represent a significant source of financing in previous studies (Birley & Westhead, 1993, p. 46; Donckels et al., 1987, p. 55). Besides that, it is notable how early the individual founders in this study had access to business angel financing. Max (P3), for example, says he had the respective angel with him "from day one", while Charly (V2) with "it came pretty quickly" and Peter (P5) with "in the very beginning" both suggest a very early point in time of angel financing. This in turn suggests that investor-entrepreneurs have good contacts to necessary sources of finance, which in turn again indicates superior social capital.

When asked about their venture financing, several founders expressed reluctance towards venture capital funding. For example, when asked if VC funding was already available, Anton (V1) and Ida (P2) respectively said: "It was already turned down for [name of venture]. I would always avoid it if possible" and "So we cancelled several term sheets because we didn’t want to have huge VCs in there". Possible reasons given are the associated information and reporting obligations (V1, V7, P2) and loss of independence (V1, V2, V7). Even though three of the founders (V2, V7 and P2) did not mention self-determination when asked about their motivation, the statements regarding venture capital funding support the assumption in the previous sub-chapter that one of the core motivations of investor-entrepreneurs is "need for independence". Whether this negative attitude towards ven-

ture capital financing is because the respective founders were previously active as investors themselves and therefore have superior information about the associated reporting obligations and the apparent loss of independence cannot be answered at this point. However, since founders with a PE background also mention it, the reason seems to be due to the desire for independence.

Considering the insights regarding human capital and the respective characteristics, I propose:

Proposition 5a: Investor-entrepreneurs, have better access to business angels as a source of funding for their venture than their novice and habitual counterparts.

Proposition 5b: Investor-entrepreneurs use savings for their foundation that they created during their time as an investor.

Proposition 5c: Investor-entrepreneurs are less likely to use venture capital financing as a result of their desire for independence.

5. Further discussion & conclusion

This study reveals several findings about investor-entrepreneurs and how they differ from their “normal” novice and habitual counterparties. Novel insights or extreme differences were condensed into propositions. Although the focus was on identifying type-specific characteristics, findings with broader regard to opportunity perceptions could also be identified. The meaning of the results, their implications, and limitations of this study and suggestions for further research are elaborated in this chapter.

5.1. The investor-entrepreneur

While previous studies (i) categorise entrepreneurs based on prior business ownership experience and (ii) consider novice founders as a homogeneous group when looking for statistical differences between the different types of entrepreneurs, the results of this study challenges both approaches. More precisely this study reveals three significant findings considering the results from the interviews and the resulting propositions. First, investor-entrepreneurs differ in multiple aspects from their "normal" novice counterparts and thus cannot be equated with other first-time founders (see Chapter 4 & Proposition 1a, 4b, 5a, b, c). For example: Investor-entrepreneurs tend to show novice as well as habitual characteristics in terms of their personal characteristics (age, gender, education). Furthermore, respective statements indicate that investor-entrepreneurs have acquired skills and knowledge such as "definition of the business model", "opportunity evaluation", or knowledge about helpful "dos and don’ts in the entrepreneurial process", that provide them with superior skills and knowledge compared to "normal" novice counterparts. This can compensate for a lack of experience as business owner, which is considered helpful in the entrepreneurial process (e.g., MacMillan, 1986; Westhead, Ucbasaran, & Wright, 2005a, pp. 396-397; Zhang, 2011, p. 5). Finally, character and behavioural differences, especially in venture financing, can be attributed to the context of the previous job as an investor. In conclusion, investor-entrepreneurs can neither be compared to novice nor habitual founders on the basis of their characteristics; however, they can by no means be considered "conventional" first-time founders.

Second, the study shows that contrary to the initial assumption, private equity investors (consisting of VC and PE) cannot be seen as a homogeneous group either (see Chapter 4 & Proposition 1b, 2a, 3a, 4a). In this study especially investor-entrepreneurs with previous VC experience represent the individuals who mainly show differences from other novice entrepreneurs and existing entrepreneur types more generally. For example, only founders with VC background tend to be significant younger than other first-time founders in terms of their age when they founded their own venture. Further, even though all founders in this study claim to have acquired necessary human capital as part of their job as an investor, there are significant differences between VC and PE in the usefulness of these skills and know-how. Similar, while most founders with a VC background say that the social network contacts and thus the social capital from their time as an investor helped them in founding their venture, only a small fraction of PE investors do so. In line with these differences in human and social capital, it is not surprising that significant differences in entrepreneurial behaviour also mostly apply only to founders with a VC background (e.g., opportunity alertness; see Chapter 4.3.1). The significant differences between these two sub-groups make it impossible to consider investor-entrepreneurs as a homogeneous group and call for a refined definition between *venture capital investor-entrepreneurs* and *private equity investor-entrepreneurs*. However, if one does so, only the venture capital investor-entrepreneur shows significant differences to existing types of entrepreneurs.

Third, this study also reveals that investor-entrepreneurs have shortcomings in operational tasks as well as their entrepreneurial mindset (Proposition 1c). So far, a comparison of this characteristic to other types of entrepreneurs is not possible since prior studies either only consider stock of experience and not task-related HC (e.g., Birley & Westhead, 1993, pp. 44-45; Westhead & Wright, 1998a, pp. 186-187) or do not take the relevant characteristics into account when examining strengths and weaknesses (e.g., Westhead, Ucbasaran, & Wright, 2005a, pp. 405-408). However, this characteristic is important for the assessment of investor-entrepreneurs and different types of entrepreneurs in general. First, it shows that previous experience (e.g., as an investor) not only brings advantages (e.g., fundraising) but also disadvantages. Second, it shows that respective shortcomings can influence entrepreneurial behaviour. For example: entrepreneurs might use their network to compensate these knowledge lacks (e.g., Leo, V5) or they might work together with other co-founders to close these gaps (cf., Westhead, Ucbasaran, & Wright, 2005b, p. 76; 91).

To integrate the investor-entrepreneur into current theoretical perspective, two different approaches exist: First, if one continues to apply the limited useful classification of en-

trepreneurs based on previous (business ownership) experience (e.g., Birley & Westhead, 1993, p. 40; Kolvereid & Bullvag, 1993, p. 276) this study calls for a more refined distinction of novice entrepreneurs on another level, similar to what is already the case for habitual entrepreneurs (i.e., serial and portfolio entrepreneurs) (e.g., Westhead & Wright, 1998a, p. 176). This refined classification allows to distinguish the private equity and venture capital investor-entrepreneur from other novice, and habitual anyway, entrepreneurs and would appropriately consider the results of this study (see Figure 3). However, the distinction of such sub-groups would be based on the previous job(s) rather than the type of exit, leading to an inconsistency in the distinction of founders and making the overall distinction of entrepreneurial types more complex.

Second, considering the results of this and shortcomings of previous studies (see Chapter 2.2.4), the established classification of entrepreneurs is unsuitable for researching differences in knowledge and skills, which in turn form the basis for entrepreneurial decisions and actions (see Appendix A). The knowledge and skills necessary for a successful foundation can be obtained through several sources such as start-up experience, previous job or general life experiences (West & Noel, 2009, pp. 6-7), as demonstrated in this study. Thus, entrepreneurs should be distinguished along their road to acquiring the "technology of entrepreneurship" (MacMillan, 1986), depending on the skills they already have and still lack, much like a to-do list for necessary resources for starting a business. Investor-entrepreneurs (and venture capital / private equity investor-entrepreneurs more accurately) with their advantages (e.g., knowledge about fundraising, dos and don'ts) and shortcomings (e.g., operational, entrepreneurial mindset) could be perfectly classified along these differentiation criteria in relation to other novice and habitual founders. However, due to the novelty of this approach as well as the absence of task-related HC and shortcomings in previous studies, a more precise positioning compared to other types of entrepreneurs is not possible at this point.

5.1.1. Additional findings on opportunity perception

In the course of studying investor-entrepreneurs, two propositions could be derived that relate to research around opportunity perception and thus to entrepreneurial behaviour more general.

First, this study suggests that venture capital funds can provide potential founders with their venture ideas by providing the necessary information and skills to identify entrepreneurial opportunities (Proposition 3b). On the one hand this supports Cooper (1985, p.84) findings that the previous employer can be an incubator and calls for an extension of existing influencing factors (geographical location, nature of business, type of company, venture size) by the variable "identify business idea". On the other hand, these findings are partly in line with earlier findings from Jarchow and Röhm (2020, p. 38) who found that intellectual property venturing funds can take on core entrepreneurial tasks such as opportunity recognition. Thereby, two things are particularly interesting: First, providing business opportunities seems to be a VC fund phenomenon only. The reason for this may again be the stage focus. VC funds, unlike PE funds for example, deal with early-stage companies and thus with innovative topics. Schumpeter (1934, pp. 66-67) argues that opportunities arise where changes in the market, technology, etc. happen. Thus, VC funds work where entrepreneurial opportunities arise. Considering that VC funds have extensive market and technology knowledge due to their investment focus, they create information asymmetries against other market participants, which makes opportunities transparent while for others not visible (Shane & Venkataraman, 2000, p. 220). Even if the opportunities still have to be identified and then exploited by the individual investors/founders themselves, VC funds act as a kind of incubator for the identification of entrepreneurial opportunities.

Second, I concluded that investor-entrepreneurs evaluate their entrepreneurial opportunities using market- and information-based valuation variables they acquired through their job as investor (Proposition 3c). This is particularly interesting as in previous studies the evaluation processes were not considered but give additional insights into the opportunity perception of the respective entrepreneurs. Further, these findings support previous research arguing that both opportunity-specific characteristics and entrepreneur-specific characteristics influence the evaluation process (e.g., Haynie et al., 2009; Welpe et al., 2012). First, „Precise evaluation of market environment" for example comprises the analysis of the competitive landscape, thus checking for rarity of the opportunity (cf., Haynie et al., 2009, p. 353). Second, with regards to the influence of entrepreneur-specific characteristics in the evaluation process, one can see that e.g., Leo (V5) clearly uses his acquired skills, knowledge, and techniques (thus HC) to evaluate his business opportunities (cf., Table 10; Appendix C).

5.2. Theoretical and practical implications

My findings contribute to two literature streams: Types of entrepreneurs and their distinction as well as on opportunity perception. Also, this study offers implications for practicians.

First and foremost, the results on investor-entrepreneurs and the refined consideration of the venture capital and private equity investor-entrepreneur challenge existing literature on types of entrepreneurs and their distinction. One the one hand, because this study proves that novice founders are not "average" (Kolvereid & Bullvag, 1993, p. 275) or lack of business ownership experience constitutes a disadvantage (Westhead, Ucbasaran, & Wright, 2005b, p. 77), thus questioning the homogeneity of first-time founders presumed over many years and studies. With the private equity and venture capital investor-entrepreneur, this study adds a first suggestion of further sub-category of novice entrepreneurs to the entrepreneurial literature. On the other hand, because the results challenge differences of entrepreneurial types along their decisions and actions

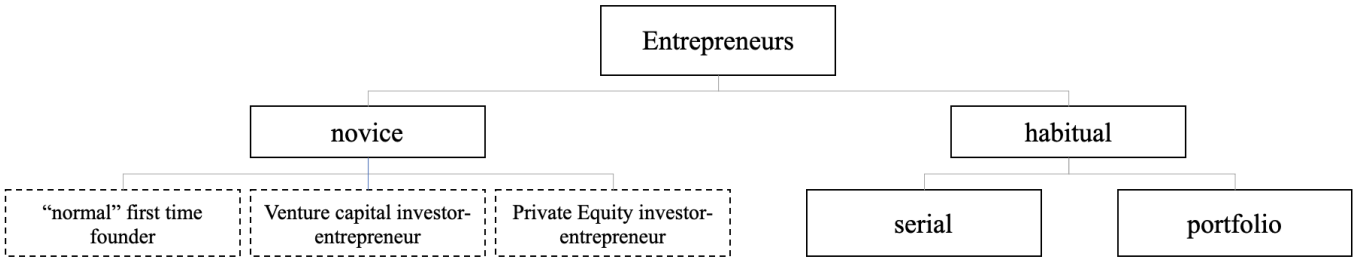

Figure 3: Entrepreneurial types (extended).

fundamentally. While academic researchers and practitioners in the area of human and social capital frequently benchmark entrepreneurial skills, knowledge and network ties against stock of prior business ownership experience (e.g., MacMillan et al., 1985, pp. 121-122; Westhead, Ucbasaran, & Wright, 2005a, pp. 395-396; Zhang, 2011, pp. 8-9) the results of this study show that relevant skills, such as those necessary for fundraising, can be acquired even without prior business ownership experience. Furthermore, I emphasize a new important consideration variable when comparing entrepreneurs along their characteristics by highlighting the importance of taking shortcomings into account. Overall, this study suggests that the availability of skills, knowledge and social ties should no longer be measured/assumed in terms of the stock of (business ownership) experience, but rather in terms of the task-relatedness of the previous job, adding to supporting conclusions from previous studies on human capital (Unger et al., 2011, pp. 26-27; West & Noel, 2009, p. 17; Westhead, Ucbasaran, & Wright, 2005a, p. 414).

Considering the results and associated implications, this study adds to existing intervariable linkages of entrepreneurial types and the relevant distinguishing variables. For illustrative purposes, I draw on the conceptual framework from Reuber and Fischer (1999, p. 31), adapted by Westhead, Ucbasaran, and Wright (2005a, p. 395) and extended by me with results from Mosey and Wright (2007) (see Figure 4).

The model comprises five new components (dashed lines). First, the investor-entrepreneur was added. Both perspectives from Chapter 5.1 have been considered: (i) The investor-entrepreneur has been included as a sub-group of novice entrepreneurs. (ii) The "road to technology of entrepreneurship" (MacMillan, 1986) was added, as proposed in Chapter 5.1. This allows the differentiation along the necessary skills and knowledge and thus a more refined distinction of diverse founders. In this context, the previous job was included, as it can impact skills and knowledge, as shown in this study. The investor-entrepreneur is positioned separately from the other novice founders because he already has important knowledge about e.g., fundraising and is therefore already more "advanced" than "normal" novice entrepreneurs. Further, as shown in this study human capital does not only comprises stock of experience, prior business ownership experience and cognitive abilities (cf., Westhead, Ucbasaran, & Wright, 2005a, p. 395) but rather includes multiple characteristics like task-related HC and individual shortcomings. Task-related HC and shortcomings represent the fourth and fifth adjustment of the model.

Second, my two side findings on entrepreneurial behaviour have broader implications on the entrepreneurial literature regarding opportunity perception (e.g., Baron & Ensley, 2006; Shane & Venkataraman, 2000; Venkataraman, 1997). First, my results and the corresponding propositions on venture capital funds show that the employer plays an important entrepreneurial role, e.g., in providing opportunities. In particular, it shows that the employer not only influences the necessary characteristics for opportunity identification (e.g., Shane, 2003, pp. 45-58), but can also be actively involved in the provision of opportunities. Second, the findings on opportunity evaluation contribute to the under-researched field of opportunity evaluation insofar as this study is one of the first to shed light on applied methods and processes rather than examining character traits of the opportunity and the entrepreneur (e.g., Haynie et al., 2009, pp. 342-346; Welpe et al., 2012, pp. 3-6). As previous studies did not address differences in opportunity evaluation between different types of entrepreneurs, this study lays the foundation for future comparisons with other types of entrepreneurs.

Lastly, the insights of this study also have implications for practitioners. First, they may influence potential future founders. This study shows that investor-, and especially venture capital investor-entrepreneurs, have advantages over their "normal" counterparts at many points necessary for setting up a venture themself. Moreover, venture capital funds seem to function as incubators for identifying entrepreneurial opportunities. Thus, similar to what some founders mentioned when asked about their motivation to become an investor (e.g., Anton (V1), David (V3) or Steven (V8)), people who feel the desire to start their own business but do not feel ready or do not have the right idea yet may decide to become a VC investor as an intermediate step to founding. Second, prior studies show that venture capital funds rely heavily on the founding team in their assessment of investment opportunities and pay particular attention to the background and relevant experience. However, studies examining VC criteria in evaluating investment opportunities do not further

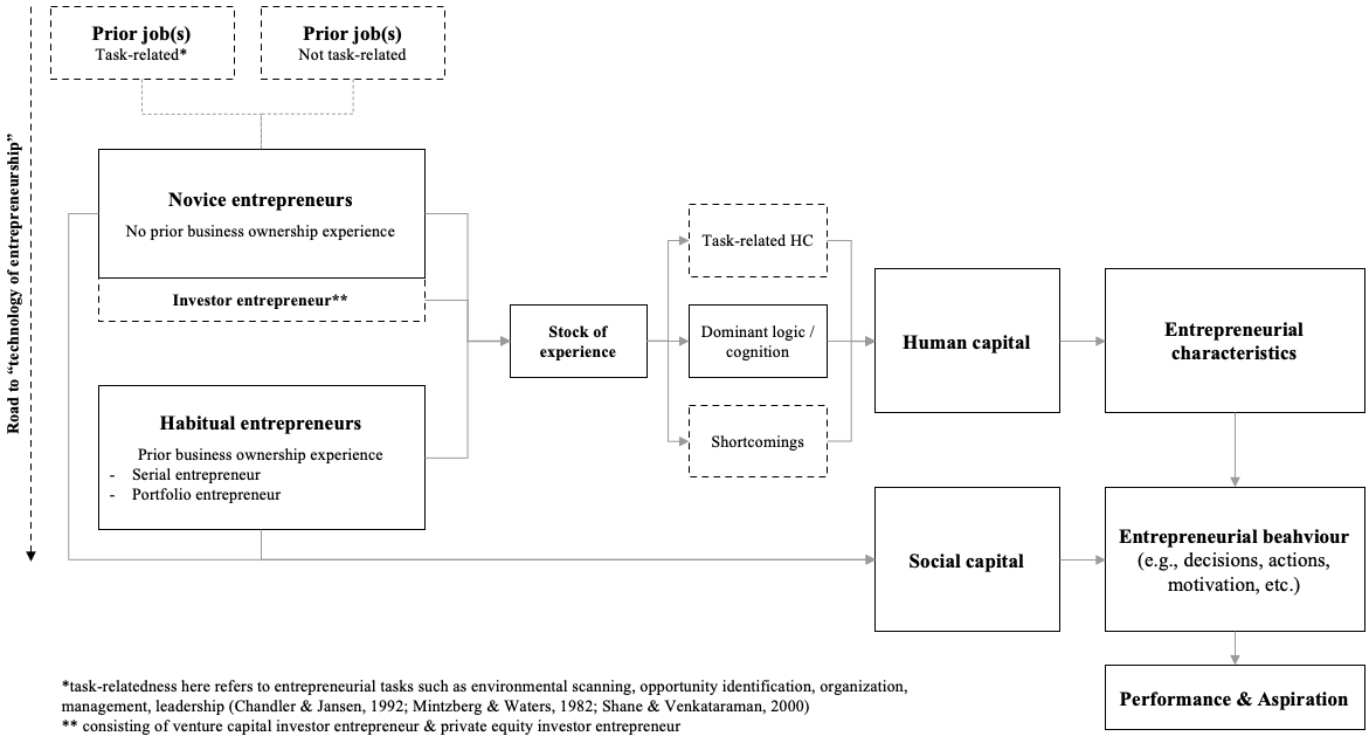

Figure 4: Visualisation of the results and extensions of existing intervariable relations (based on Westhead, Ucbasaran, and Wright (2005b)).

specify "background/experience" (J. Hall & Hofer, 1993, p. 27) or "track record relevant to venture" (MacMillan et al., 1985, p. 121), thus encompassing the complete pool of available human capital. The results of this study may encourage practitioners to focus more on the task-relatedness of experience and thus the contextual requirements when evaluating founding teams in the future and thus to create more specified evaluation processes.

5.3. Limitations and further research

Before I conclude, it should be noted that this study is not free of limitations. First, this study uses a cross-sectional design to explore the research question. Therefore, the findings only provide a snapshot of the current situation. Moreover, the focus was primarily on the foundation and early-stage of the respective ventures. A longitudinal study would provide a better understanding of the characteristics, shortcomings, their relevance, and impact on mid- and long-term venture creation.

Second, the study is limited to 13 cases and only considers founders in Germany that were investigated. Thus, as for much qualitative research, the results are limited in their generalizability. Further research should therefore test the proposed definition of the investor-entrepreneur and its refined sub-groups as well as the newly introduced relationships regarding founder-related differentiation variables in a quantitative study consisting of a broader range of cases. In selecting the sample, they should consider more geographical diversity. In addition, a quantitative research method should be chosen, in order to enable statistical rather than qualitative generalisability (Eisenhardt, 1989, p. 547).

Third, this study does not explore cognitive differences because of the limited scope and time available for this study and the challenge of exploring these attributes qualitatively. However, cognitive abilities represent an important component in the entrepreneurial literature (e.g., Alvarez & Busenitz, 2001, p. 764; Shane, 2003, p. 54) and more specifically in distinguishing different types of entrepreneurs (Westhead, Ucbasaran, & Wright, 2005b, pp. 86-89). Therefore, future research should take these into account when validating and comparing venture capital/private equity investor-entrepreneurs and investor-entrepreneurs more broadly.

Fourth, the transcripts were submitted to the respective interview participants for approval. Thereby, multiple interviewees returned the transcript adjusted. As a result, some initial statements were changed and thus distorted, which could have led to deleted important findings and thus not being considered in the analysis.

Finally, although much is already known about different traits and characteristics that are helpful for entrepreneurship (e.g., Ozgen & Baron, 2007; Shane, 2003; Unger et al., 2011), to the best of my knowledge there is no summary literature of the influencing variables. Therefore, further research should explore and define the proposed to-do list (Chapter 5.1) consisting of necessary skills and knowl-

edge for the "technology of entrepreneurship" (MacMillan, 1986). Besides that, this study provides the basis for further research around a refined definition of novice entrepreneurs. In addition to a proposed possible differentiation approach and a first suggested sub-group (venture capital and private equity investor-entrepreneur), interview partners like Robby (V7), Felix (P1) or Max (P3) mention founders with previous consulting experience and their characteristics and shortcomings, thus already suggesting another possible subgroup of novice entrepreneurs. For example, Max (P3) says: "Accordingly, you are definitely more likely to have consultants who have taken the path [of becoming an entrepreneur] because, depending on the consultancy, that is also your mindset; they are a dime a dozen." A further research question would therefore be: *How do consulting entrepreneurs differ from other (e.g., "normal" or investor) entrepreneurs?* In addition to this sub-group, further research should look at which professional/industrial fields parallel helpful entrepreneurial tasks and thus possible task-related HC can be obtained. The corresponding founders could form further entrepreneurial groups with significant entrepreneurial differences.

5.4. Conclusion

This explorative study aimed to compare the investor-entrepreneur with existing types of entrepreneurs (i.e., novice and habitual) regarding their personal and entrepreneurial characteristics. Guided by the research question: "*How do investor-entrepreneurs differ from novice and habitual entrepreneurs?*" I interviewed 13 founders from Germany with previous experience in venture capital and private equity. The results of this study reveal that first investor-entrepreneurs differ from previously defined novice entrepreneurs in many ways. More specifically, they show characteristics of novice and habitual entrepreneurs as well as entirely new traits. Second, investor-entrepreneurs (consisting of VC and PE investors) cannot be seen as a homogeneous group and third, besides some superior knowledge (e.g., fundraising), they also have shortcomings in the areas of personal management and entrepreneurial mindset. Considering these results, investor-entrepreneurs cannot be assigned to existing types of entrepreneurs. Thus, to integrate the investor-entrepreneur into current theoretical perspective, this study either suggests a more refined differentiation of first-time founders, in contrast to the so far homogeneous approach in previous studies or suggests a distinction of entrepreneurs along necessary skills they already have and still lack, following a resource-based view. I hope these findings motivate future research to adapt the suggested novel differentiation approach and investigate the investor-entrepreneur as well as other types of fist-time founder (e.g., consulting-entrepreneur) in more depth.

References

Adler, P. S., & Kwon, S.-W. (2002). Social capital: Prospects for a new concept. *Academy of Management Review*, *27*(1), 17–40.

Aldrich, H., & Zimmer, C. (1986). Entrepreneurship through social networks. *The Art and Science of Entrepreneurship*, *3*, 23.

Alvarez, S. A., & Busenitz, L. W. (2001). The entrepreneurship of resource-based theory. *Journal of Management*, *27*(6), 755–775.

Baron, R. A. (2006). Opportunity recognition as pattern recognition: How entrepreneurs "connect the dots" to identify new business opportunities. *Academy of Management Perspectives*, *20*(1), 104–119.

Baron, R. A., & Ensley, M. D. (2006). Opportunity recognition as the detection of meaningful patterns: Evidence from comparisons of novice and experienced entrepreneurs. *Management Science*, *52*(9), 1331–1344.

Becker, G. S. (1964). *Human capital: A theoretical and empirical analysis, with special reference to education*. University of Chicago Press.

Becker, G. S. (1993). Nobel lecture: The economic way of looking at behavior. *Journal of Political Economy*, *101*(3), 385–409.

Bell, E., Bryman, A., & Harley, B. (2018). *Business Research Methods*. Oxford University Press.

Berk, J., & DeMarzo, P. (2016). *Corporate Finance, GE*. Pearson Australia Pty Limited.

Birley, S. (1985). The role of networks in the entrepreneurial process. *Journal of Business Venturing*, *1*(1), 107–117.

Birley, S., & Westhead, P. (1993). A comparison of new businesses established by'novice'and'habitual'founders in Great Britain. *International Small Business Journal*, *12*(1), 38–60.

Block, J., Fisch, C., Vismara, S., & Andres, R. (2019). Private equity investment criteria: An experimental conjoint analysis of venture capital, business angels, and family offices. *Journal of Corporate Finance*, *58*, 329–352.

Bryman, A. (2006). Integrating quantitative and qualitative research: how is it done? *Qualitative Research*, *6*(1), 97–113.

Brüderl, J., & Preisendörfer, P. (1998). Network support and the success of newly founded business. *Small Business Economics*, *10*(3), 213–225.

Carbonara, E., Tran, H. T., & Santarelli, E. (2019). Determinants of novice, portfolio, and serial entrepreneurship: an occupational choice approach. *Small Business Economics*, 1–29.

Cassala, C. (2018, 4. August 2021). *Noch immer in der Minderheit*. Retrieved from https://www.vc-magazin.de/blog/2018/12/11/noch-immer-in-der-minderheit/2/

Chandler, G. N., & Jansen, E. (1992). The founder's self-assessed competence and venture performance. *Journal of Business Venturing*, *7*(3), 223–236.

Charles, V., & Gherman, T. (2013). Factors influencing peruvian women to become entrepreneurs. *World Applied Sciences Journal*, *27*(10), 1345–1354.

Chong, H. G. (2008). Measuring performance of small-and-medium sized enterprises: The grounded theory approach. *Journal of Business & Public Affairs*.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, *94*, S95–S120.

Cooper, A. C. (1985). The role of incubator organizations in the founding of growth-oriented firms. *Journal of Business Venturing*, *1*(1), 75–86.

Cooper, A. C., Gimeno-Gascon, F. J., & Woo, C. Y. (1994). Initial human and financial capital as predictors of new venture performance. *Journal of Business Venturing*, *9*(5), 371–395.

Corbin, J., & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). Sage Publications.

Cumming, D. (2012). *The Oxford Handbook of Private Equity*. Oxford University Press.

Donckels, R., Dupont, B., & Michel, P. (1987). Multiple Business Starters Who? Why? What? *Journal of Small Business & Entrepreneurship*, *5*(1), 48–63.

Dubini, P., & Aldrich, H. (1991). Personal and extended networks are central to the entrepreneurial process. *Journal of Business Venturing*, *6*, 305.

Durst, S., & Wilhelm, S. (2012). Knowledge management and succession planning in SMEs. *Journal of Knowledge Management*.

Edmondson, A. C., & McManus, S. E. (2007). Methodological fit in management field research. *Academy of Management Review*, *32*(4), 1246–1264.

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, *14*(4), 532–550.

Elfring, T., & Hulsink, W. (2003). Networks in entrepreneurship: The case of high-technology firms. *Small Business Economics*, *21*(4), 409–422.

Fried, V. H., & Hisrich, R. D. (1994). Toward a model of venture capital investment decision making. *Financial Management*, 28–37.

Glaser, B. G., & Strauss, A. L. (2017). *Discovery of grounded theory: Strategies for qualitative research*. Routledge.

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, *78*(6), 1360–1380.

Grant, R. M. (1991). The resource-based theory of competitive advantage: implications for strategy formulation. *California Management Review*, *33*(3), 114–135.

Gruber, M., & MacMillan, I. C. (2017). Entrepreneurial behavior: A reconceptualization and extension based on identity theory. *Strategic Entrepreneurship Journal*, *11*(3), 271–286.

Hall, J., & Hofer, C. W. (1993). Venture capitalists' decision criteria in new venture evaluation. *Journal of Business Venturing*, *8*(1), 25–42.

Hall, P. (1995). Habitual owners of small businesses. In *Small Firms: Partnership of Growth* (pp. 217–230). Paul Chapman Publishing.

Harris, R. S., Jenkinson, T., & Kaplan, S. N. (2014). Private equity performance: What do we know? *The Journal of Finance*, *69*(5), 1851–1882.

Haynie, J. M., Shepherd, D. A., & McMullen, J. S. (2009). An opportunity for me? the role of resources in opportunity evaluation decisions. *Journal of Management Studies*, *46*(3), 337–361.

Hochberg, Y. V., Ljungqvist, A., & Lu, Y. (2007). Whom you know matters: Venture capital networks and investment performance. *The Journal of Finance*, *62*(1), 251–301.

Jarchow, S., & Röhm, A. (2020). Business builders, contractors, and entrepreneurs—An exploratory study of IP venturing funds. *Journal of Small Business Management*, 1–46.

Kirzner, I. M. (1973). *Competition and entrepreneurship*. University of Chicago Press.

Kolvereid, L., & Bullvag, E. (1993). Novices versus experienced business founders: An exploratory investigation. *Entrepreneurship Research: Global Perspectives*, 275–285.

MacMillan, I. C. (1986). To really learn about entrepreneurship, let's study habitual entrepreneurs. *Journal of Business Venturing*, *1*, 241–243.

MacMillan, I. C., Siegel, R., & Narasimha, P. S. (1985). Criteria used by venture capitalists to evaluate new venture proposals. *Journal of Business Venturing*, *1*(1), 119–128.

McGrath, R. G., & MacMillan, I. C. (2000). *The entrepreneurial mindset: Strategies for continuously creating opportunity in an age of uncertainty* (Vol. 284). Harvard Business Press.

Merz, C. (2008). *Erfahrene Unternehmensgründer*. Springer-Verlag.

Metrick, A., & Yasuda, A. (2010). The economics of private equity funds. *Review of Financial Studies*, *23*(6), 2303–2341.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage.

Mosey, S., & Wright, M. (2007). From human capital to social capital: A longitudinal study of technology–based academic entrepreneurs. *Entrepreneurship Theory and Practice*, *31*(6), 909–935.

Ozgen, E., & Baron, R. A. (2007). Social sources of information in opportunity recognition: Effects of mentors, industry networks, and professional forums. *Journal of Business Venturing*, *22*(2), 174–192.

Perry, C., & Jensen, O. (2001). Approaches to combining induction and deduction in one research study. In *Conference of the australian and new zealand marketing academy, auckland, new zealand.*

Plehn-Dujowich, J. (2010). A theory of serial entrepreneurship. *Small Business Economics*, *35*(4), 377–398.

Politis, D. (2008). Does prior start-up experience matter for entrepreneurs' learning? A comparison between novice and habitual entrepreneurs. *Journal of Small Business and Enterprise Development*.

Preller, R., Patzelt, H., & Breugst, N. (2020). Entrepreneurial visions in founding teams: Conceptualization, emergence, and effects on opportunity development. *Journal of Business Venturing*, *35*(2), 105914.

Retterath, A., & Braun, R. (2020). Benchmarking Venture Capital Databases. *Available at SSRN*.

Reuber, A. R., & Fischer, E. (1999). Understanding the consequences of founders' experience. *Journal of Small Business Management*, *37*(2), 30.

Robbie, K., & Wright, M. (1996). *Management buy-ins: entrepreneurship, active investors, and corporate restructuring*. Manchester University Press.

Sandberg, W. R., & Hofer, C. W. (1987). Improving new venture performance: The role of strategy, industry structure, and the entrepreneur. *Journal of Business Venturing*, *2*(1), 5–28.

Schumpeter, J. A. (1934). *The Theory of Economic Development*. Harvard University Press.

Shane, S. (2003). *A general theory of entrepreneurship: The individual-opportunity nexus*. Edward Elgar Publishing.

Shane, S., & Cable, D. (2002). Network ties, reputation, and the financing of new ventures. *Management Science*, *48*(3), 364–381.

Shane, S., & Venkataraman, S. (2000). The promise of entrepreneurship as a field of research. *Academy of Management Review*, *25*(1), 217–226.

Smith, N. R. (1967). The entrepreneur and his firm: The relationship between type of man and type of company. *Occasional papers, Bureau of Business and Economic research, Michigan State University*, *109*.

Stam, W., Arzlanian, S., & Elfring, T. (2014). Social capital of entrepreneurs and small firm performance: A meta-analysis of contextual and methodological moderators. *Journal of Business Venturing*, *29*(1), 152–173.

St-Jean, E., & Audet, J. (2012). The role of mentoring in the learning development of the novice entrepreneur. *International Entrepreneurship and Management Journal*, *8*(1), 119–140.

Suddaby, R. (2006). *From the editors: What grounded theory is not.* Academy of Management Briarcliff Manor, NY 10510.

Teten, D., & Farmer, C. (2010). Where are the deals? private equity and venture capital funds' best practices in sourcing new investments. *The Journal of Private Equity*, *14*(1), 32–52.

Unger, J. M., Rauch, A., Frese, M., & Rosenbusch, N. (2011). Human capital and entrepreneurial success: A meta-analytical review. *Journal of Business Venturing*, *26*(3), 341–358.

Uzzi, B. (1997). Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative Science Quarterly*, *42*(1), 213–241.

Uzzi, B. (1999). Embeddedness in the making of financial capital: How social relations and networks benefit firms seeking financing. *American Sociological Review*, 481–505.

Venkataraman, S. (1997). The distinctive domain of entrepreneurship research. In *Seminal ideas for the next twenty-five years of advances.* Emerald Publishing Limited.

Vollstedt, M., & Rezat, S. (2019). An introduction to grounded theory with a special focus on axial coding and the coding paradigm. *Compendium for early Career Researchers in Mathematics Education*, *13*, 81–100.

Welpe, I. M., Spörrle, M., Grichnik, D., Michl, T., & Audretsch, D. B. (2012). Emotions and opportunities: The interplay of opportunity evaluation, fear, joy, and anger as antecedent of entrepreneurial exploitation. *Entrepreneurship Theory and Practice*, *36*(1), 69–96.

West, G. P., & Noel, T. W. (2009). The impact of knowledge resources on new venture performance. *Journal of Small Business Management*, *47*(1), 1–22.

Westhead, P., Ucbasaran, D., & Wright, M. (2005a). Decisions, actions, and performance: do novice, serial, and portfolio entrepreneurs differ? *Journal of Small Business Management*, *43*(4), 393–417.

Westhead, P., Ucbasaran, D., & Wright, M. (2005b). Experience and cognition: do novice, serial and portfolio entrepreneurs differ? *International Small Business Journal*, *23*(1), 72–98.

Westhead, P., Ucbasaran, D., Wright, M., & Binks, M. (2005). Novice, serial and portfolio entrepreneur behaviour and contributions. *Small Business Economics*, *25*(2), 109–132.

Westhead, P., & Wright, M. (1998a). Novice, portfolio, and serial founders: are they different? *Journal of Business Venturing*, *13*(3), 173–204.

Westhead, P., & Wright, M. (1998b). Novice, portfolio, and serial founders in rural and urban areas. *Entrepreneurship theory and practice*, *22*(4), 63–100.

Wiklund, J., & Shepherd, D. (2003). Knowledge-based resources, entrepreneurial orientation, and the performance of small and medium-sized businesses. *Strategic Management Journal*, *24*(13), 1307–1314.

Witt, P. (2004). Entrepreneurs' networks and the success of start-ups. *Entrepreneurship & Regional Development*, *16*(5), 391–412.

Wright, M., & Robbie, K. (1998). Venture capital and private equity: A review and synthesis. *Journal of Business Finance & Accounting*, *25*(5-6), 521–570.

Yin, R. K. (2009). *Case study research: Design and methods* (Vol. 5). Sage.

Zhang, J. (2011). The advantage of experienced start-up founders in venture capital acquisition: evidence from serial entrepreneurs. *Small Business Economics*, *36*(2), 187–208.

Junior Management Science 8(1) (2023) 68-95

Junior Management Science

journal homepage: www.jums.academy

The role of management control systems in reducing and preventing managerial misreporting: A critical analysis

Die Bedeutung der Unternehmenssteuerung für die Sicherstellung einer wahrheitsgemäßen Berichterstattung: Eine kritische Analyse

Michael Keck

Heinrich-Heine-Universität Düsseldorf

Abstract

To manage the company efficiently, management control systems depend on truthful information from their employees. In contrast to the classical principal-agent theory, in which managers represent purely self-interested individuals, behavioral science research has shown that honesty preferences play a role in managers' decisions. The aim of this thesis is to investigate how management control systems can contribute to reducing and preventing managerial misreporting through a clever design of control instruments. The object-of-control framework of Merchant and Van der Stede (Merchant & Van der Stede, 2017) is used in the context of this study. Based on literature, formal and informal control instruments are examined for their potential to bring truthful reporting in the interest of the company. The results show that management control systems can contribute to reducing and preventing managerial misreporting through a clever design of the control instruments, and that companies can benefit from managers' honesty preferences. The management control system should take care to behave fairly toward managers itself and, in particular, implement measures that improve social norms for honesty in the company, thereby creating a mutual-monitoring effect. Since control instruments can also have a negative impact on truthful reporting if they are poorly designed, companies should ensure that their management control system is holistically designed.

Zusammenfassung

Zur effizienten Steuerung des Unternehmens ist die Unternehmensführung auf wahrheitsgemäße Informationen ihrer Manager angewiesen. Anders als in der klassischen Prinzipal-Agenten-Theorie, in der Manager rein eigennutzorientierte Individuen darstellen, hat die verhaltenswissenschaftliche Forschung gezeigt, dass Ehrlichkeitspräferenzen in den Entscheidungen von Managern eine Rolle spielen. Das Ziel der vorliegenden Arbeit ist es, zu untersuchen, wie die Unternehmensführung durch eine geschickte Ausgestaltung von Steuerungsinstrumenten zu einer wahrheitsgemäßen Berichterstattung beitragen kann. Den Rahmen für die Untersuchung bietet das object-of-control-Framework von Merchant und Van der Stede (2017). Literaturbasiert werden formelle und informelle Steuerungsinstrumente auf deren Potential hin untersucht, eine wahrheitsgemäße Berichterstattung im Sinne des Unternehmens herbeizuführen. Die Ergebnisse zeigen, dass die Unternehmenssteuerung durch eine geschickte Ausgestaltung der Steuerungsinstrumente von den Ehrlichkeitspräferenzen der Manager profitieren und zu einer wahrheitsgemäßen Berichterstattung beitragen kann. Die Unternehmensführung sollte darauf achten sich selbst fair gegenüber Managern zu verhalten und insbesondere Maßnahmen umsetzen, die soziale Normen zur Ehrlichkeit im Unternehmen verbessern und dadurch einen Mutual-Monitoring-Effekt bewirken. Da Steuerungsinstrumente bei einer mangelhaften Ausgestaltung auch einen negativen Einfluss auf die wahrheitsgemäße Berichterstattung nehmen können, sollten Unternehmen auf eine ganzheitliche Ausgestaltung ihres Management Control Systems achten.

Keywords: Wahrheitsgemäße Berichterstattung; Management Control System; Ehrlichkeit; Offenlegung; formelle und informelle Steuerungsinstrumente.

DOI: https://doi.org/10.5282/jums/v8i1pp68-95

1. Einleitung

> „Honesty is the best policy – when there is money in it" (Twain, 1998, S. 33)

Menschen stellen sich gerne hinter die Idee, dass es die beste Strategie ist, die Wahrheit zu sagen und Menschen auch grundsätzlich dazu neigen, ehrlich zu sein (vgl. Salterio & Webb, 2006, S. 919). In der Wirtschaftswelt lassen große Skandale wie der Bilanzbetrug bei Wirecard oder der Volkswagen-Abgasskandal jedoch Zweifel daran aufkommen. Auch im Unternehmen scheint Ehrlichkeit nicht am längsten zu währen. So berichten in einer Studie von Libby und Lindsay (2010) 77 Prozent der befragten kanadischen Teilnehmer (und sogar 86 Prozent der US-Teilnehmer), dass sie in den letzten beiden Jahren bewusst einfachere Ziele verhandelt hätten, als sie für die Studienteilnehmer realistisch erschienen (vgl. Libby & Lindsay, 2010, S. 65).

Dabei ist gerade die Unternehmensführung auf wahrheitsgemäße Informationen ihrer Mitarbeiter und Manager angewiesen, um das Unternehmen effizient steuern zu können (vgl. Brunner & Ostermaier, 2019a, S. 127). Manager verfügen im Unternehmen über private Informationen. Unternehmen wollen diese Information für die Erreichung ihrer Ziele nutzen, indem sie Manager bei ihren Entscheidungen miteinbeziehen (vgl. Lohmann, 2013, S. 56). Dabei müssen die Ziele des Managers jedoch nicht notwendigerweise mit den Zielen des Unternehmens übereinstimmen (vgl. Brown, Fisher, Peffer & Sprinkle, 2017, S. 31). Bei abweichenden Zielen hat der Manager einen Anreiz seine eigenen Informationen zurückzuhalten oder falsch zu berichten, um dadurch einen Vorteil zu erzielen (vgl. Hannan, Rankin & Towry, 2006, S. 892). Dadurch ergibt sich ein Anreizproblem, das es im Unternehmen zu lösen gilt (vgl. Chong & Wang, 2019, S. 275).

In den Wirtschaftswissenschaften ist die Sichtweise des Homo oeconomicus vorherrschend (vgl. Gneezy, 2005, S. 384). Demnach sind Individuen rein eigennutzorientiert und streben danach ihren eigenen Nutzen zu maximieren (vgl. Baiman, 1990, S. 342). Die Anreizproblematik kommt durch die Aufgaben- und Entscheidungsdelegation in Unternehmen zustande und wird in der in der Prinzipal-Agenten-Theorie durch die Ausgestaltung von Verträgen gelöst (vgl. Ebers & Gotsch, 2019, S. 207 f.). Ohne Vorliegen eines solchen Vertrages würden Agenten niemals wahrheitsgemäß berichten, wenn ihnen dadurch ein eigener Nutzen entgeht (vgl. Rankin, Schwartz & Young, 2008, S.1084). Ehrlichkeitspräferenzen spielen in der klassischen Agency-Theorie folglich keine Rolle (vgl. Paz, Reichert & Woods, 2014, S. 89).

Um zu untersuchen, ob und in welchem Ausmaß Manager ihre privaten Informationen wahrheitsgemäß berichten, kann betrachtet werden, wann eine Person lügt. Um eine Lüge handelt es sich, wenn eine Person aktiv eine Aussage tätigt und sich zum Zeitpunkt der Aussage darüber bewusst ist, dass diese Aussage falsch ist. Grund dafür ist die Möglichkeit einen Nutzen daraus erzielen zu können (vgl. Grover, 2005, S. 148 f.). Je geringer die Chance ist, dass die Lüge aufgedeckt wird, desto wahrscheinlicher ist es, dass eine Person lügt (vgl. Grover, 2005, S. 150). Aus einer ökonomischen Perspektive lässt sich Ehrlichkeit definieren als der Anteil der vorhandenen privaten Informationen, der kommuniziert wird (vgl. Hannan et al., 2006, S. 889).

Die Gestaltung von Verträgen zwischen Prinzipal und Agent erfolgt demzufolge mit dem Ziel, den Agenten zu einer wahrheitsgemäßen Berichterstattung zu motivieren (vgl. Weiskirchner-Merten, 2020, S. 24 f.). Die verhaltenswissenschaftliche Forschung hat sich in den letzten Jahren in einer Reihe von Studien mit den Ehrlichkeitspräferenzen von Managern beschäftigt und ist zu dem Ergebnis gekommen, dass diese das Reporting-Verhalten von Managern beeinflussen, obwohl sie nicht durch die klassische Prinzipal-Agenten-Theorie zu erklären sind (vgl. hierzu u. a. Evans, Hannan, Krishnan & Moser, 2001; Rankin et al., 2008; Douthit & Stevens, 2015). Unternehmen sollten sich entsprechend damit befassen, wie sie das Verhalten ihrer Manager steuern können, um von Ehrlichkeitspräferenzen zu profitieren (vgl. Evans et al., 2001, S. 553).

Um Verhalten und Entscheidungen der Mitarbeiter im Sinne des Unternehmens zu steuern und diese auf die Ziele des Unternehmens auszurichten, greifen Unternehmen auf eine Vielzahl von Steuerungsinstrumenten (im Laufe der Arbeit auch als Control-Instrumente bezeichnet) zurück. Die Gesamtheit dieser Instrumente wird als Management Control System (MCS) bezeichnet (vgl. Merchant & Van der Stede, 2017, S. 8). Die Definition von MCS unterlag dabei im Zeitablauf vielen Veränderungen und ist aufgrund zahlreicher Einflüsse schwierig einheitlich zu formulieren (vgl. Chenhall, 2003, S. 129). Vielmehr haben sich in der Theorie verschiedene Ansätze herausgebildet, wie etwa das levers-of-control-Framework nach Simons (1995) oder der object-of-control-Ansatz von Merchant und Van der Stede (2017). Der object-of-control-Ansatz stellt dabei menschliches Verhalten in den Vordergrund und verfolgt das Ziel, unerwünschtes Verhalten im Unternehmen zu vermeiden und zu wünschenswertem Verhalten zu ermutigen (vgl. Merchant & Van der Stede, 2017, S. 11).

Control-Instrumente leisten so einen Beitrag dazu, dass Manager ihre Zeit darauf verwenden, die Präferenzen des Unternehmens zu verfolgen, wozu eben auch die Offenlegung privater Informationen zählt (vgl. Covaleski, Evans III, Luft & Shields, 2003, S. 12). Steuerungsinstrumente stellen dabei keine unter allen Rahmenbedingungen optimalen Handlungsempfehlungen dar. Vielmehr wird ihre Wirksamkeit durch eine Vielzahl von Einflussfaktoren wie z. B. der Unternehmensgröße oder der Technologienutzung im Unternehmen beeinflusst (vgl. Chenhall, 2003, S. 160 f.). Ein gutes MCS wird demnach vielmehr durch das zugrunde liegende Unternehmen definiert und ist abhängig von der Situation im Unternehmen (vgl. Otley, 1980, S. 425). Einzelne Steuerungsinstrumente können untereinander korreliert sein, was eine ganzheitliche Betrachtung des MCS notwendig macht (vgl. Fisher, 1998, S. 61). Dadurch können Unternehmen nachvollziehen wie sich Control-Instrumente zueinander verhalten und dadurch zur Erreichung der Ziele beitragen (vgl. Malmi & Brown, 2008, S. 288; Milgrom & Roberts, 1995,

S. 190). Zur Implementierung von effektiven Steuerungsinstrumenten kann das Controlling eines Unternehmens einen wichtigen Beitrag leisten (vgl. Schäffer, 2013, S. 300).

Die vorliegende Arbeit untersucht, wie die Unternehmensführung durch eine geschickte Ausgestaltung von Steuerungsinstrumenten zu einer wahrheitsgemäßen Berichterstattung beitragen kann. Der Fokus der Arbeit liegt dabei auf der internen Kommunikation zwischen der Unternehmensführung und Managern, die über einen Informationsvorsprung verfügen, wie es etwa in der Budgetierung oder der Investitionssteuerung der Fall ist. Den Rahmen für die Untersuchung bietet das object-of-control-Framework von Merchant und Van der Stede (2017).

Im zweiten Kapitel werden zunächst die Grundlagen zur wahrheitsgemäßen Berichterstattung vermittelt, bevor auf Budgetary Slack sowie das object-of-control-Framework eingegangen wird. Kapitel 3 beschäftigt sich mit der Ausgestaltung formeller Controls. Dazu zählen Results Controls und Action Controls. Anschließend werden in Kapitel 4 informelle Controls in Form von Personnel Controls und Cultural Controls diskutiert. Die Arbeit schließt mit einem Fazit.

2. Ursachen für das Auftreten von Steuerungsproblemen vor dem Hintergrund einer wahrheitsgemäßen Berichterstattung

2.1. Ehrlichkeit in der Prinzipal-Agenten-Theorie

Die Agency-Theorie geht auf die beiden Wirtschaftswissenschaftler Jensen und Meckling (1976) zurück. Sie definieren eine Agency-Beziehung als einen Vertrag zwischen ein oder mehreren Personen (Prinzipal) mit einer anderen Person (Agent). Im Rahmen dieses Vertrages soll der Agent Leistungen im Sinne des Prinzipals erbringen, was die Delegation von Entscheidungsrechten an diesen erfordert (vgl. Jensen & Meckling, 1976, S. 308). In der Prinzipal-Agenten-Theorie wird angenommen, dass sowohl Prinzipal als auch Agent rationale Individuen sind, die mit ihren Entscheidungen das Ziel verfolgen ihren eigenen Nutzen zu maximieren. Dabei strebt der Agent nach einem Zuwachs seines Wohlstandes, unterliegt gleichzeitig einem Arbeitsleid (d. h. er bevorzugt weniger Arbeit gegenüber mehr Arbeit) und weist eine Risikoneigung auf. Beide Parteien kennen die Präferenz zur Eigennutzmaximierung der jeweils anderen Partei und unterstellen, dass die andere Partei danach strebt diese Präferenz durchzusetzen (vgl. Covaleski et al., 2003, S. 11 f.). Zentral in der Agency-Theorie ist das Vorhandensein von Informationsasymmetrien, wobei der Agent über mehr entscheidungsrelevante Informationen verfügt als der Prinzipal. Im Rahmen des Vertrages greift der Prinzipal auf Anreizmechanismen zurück, um die bestehende Informationsasymmetrie abzubauen (vgl. Blum, Dudley, Leibbrand & Weiske, 2005, S. 155). Mithilfe der Anreizmechanismen versucht der Prinzipal somit Probleme, die sich in der Agency-Theorie ergeben, zu beseitigen. Obwohl in der Theorie nicht einheitlich definiert, können diese in die vier folgenden Problemtypen unterteilt werden: Hidden Characteristics (Problem der Auswahl eines geeigneten Agenten), Hidden Intention (Problem die Absichten des Agenten nicht beobachten zu können), Hidden Knowledge (Probleme, die sich aus dem Wissensvorsprung des Agenten ergeben können) und Hidden Action (Prinzipal sieht lediglich das Ergebnis, nicht aber die Handlungen des Agenten) (vgl. Ebers & Gotsch, 2019, S. 212 f.). Durch die Beseitigung der Informationsasymmetrien entstehen dem Prinzipal Kosten. Ziel der Prinzipal-Agenten-Theorie ist es, diese Kosten zu minimieren (vgl. Ebers & Gotsch, 2019, S. 211). Demnach wird sich der Agency-Vertrag mit den geringsten Kosten als der am effizientesten ausgestaltete Vertrag gegenüber anderen Verträgen durchsetzen (vgl. Jensen, 1983, S. 331).

Der Besitz privater Informationen stellt für den Agenten in der Agency-Theorie folglich einen elementaren Vorteil dar. Der Agent befindet sich in einer umso besseren Verhandlungsposition, je stärker der Grad an asymmetrischer Informationsverteilung ausgeprägt ist und je mehr private Informationen der Agent damit besitzt. Zu den Möglichkeiten, die sich für den Agenten ergeben, zählt auch unethisches Verhalten, wie etwa die nicht wahrheitsgemäße Weitergabe seiner privaten Informationen (vgl. Chong & Wang, 2019, S. 279).

Aufgrund der Annahme, dass Agenten Nutzenmaximierer sind, werden Agenten diese Möglichkeiten nutzen, wenn sich dadurch ein Vorteil für sie ergibt (vgl. Schwering, 2016, S. 86). Ob und inwiefern soziale Präferenzen eine Rolle in der Nutzenfunktion des Agenten spielen, wird dabei in der Agency-Sichtweise häufig vernachlässigt (vgl. Dees, 1992, S. 34).

Unterstützung für eine nicht notwendige Betrachtung der Wirkung sozialer Präferenzen findet sich in Studien, in denen lediglich zehn Prozent der Menschen über sich selbst berichten, dass sie sich in allen denkbaren Situationen ehrlich verhalten würden. Passend dazu lässt sich feststellen, dass oftmals bereits ein kleiner und nicht notwendigerweise finanzieller Anreiz ausreicht, um Menschen zum Lügen zu motivieren (vgl. Salterio & Webb, 2006, S. 920 f.). Auch laut Nagin, Rebitzer, Sanders und Taylor (2002) besitzt diese klassische Agency-Betrachtung durchaus seine Berechtigung. In einem Experiment reagierte ein Großteil der Teilnehmer auf den Wegfall von Monitoring-Maßnahmen mit einem opportunistischen Verhalten, was bedeutet, dass sie ihren persönlichen Nutzen erhöhten (vgl. Nagin et al., 2002, S. 870).

Unter optimalen Bedingungen erwarten Wirtschaftswissenschaftler, dass annähernd jedes Individuum bereit ist zu lügen, wenn es dadurch einen Vorteil erlangt. Die Bedingungen dafür sind eine komplette Anonymität im Sinne fehlender Überwachung, das Auftreten von keinerlei Reputationseffekten sowie keinerlei persönliche Verantwortung. Selbst unter Vorliegen dieser Bedingungen zeigt sich in der Forschung jedoch, dass sich nicht alle Individuen zu einem unethischen Verhalten entscheiden (vgl. Salterio & Webb, 2006, S. 920). Bereits Baiman und Lewis (1989) untersuchten im Rahmen der Agency-Theorie die Ehrlichkeit der Agenten. Sie errechneten eine Hemmschwelle von 0,25 Dollar. Erhalten die Agenten aus einer Lüge eine größere Entschädigung, so würden sie sich gegen Ehrlichkeit entscheiden. Aufgrund des

geringen Betrages sahen sie darin die Annahme des Agency-Modells bestätigt, wonach die Präferenz zur Ehrlichkeit von Individuen so gering ist, dass sie vernachlässigt werden könne (Baiman & Lewis, 1989, S. 14 f.).

Überlegungen, wie sich ethische Ausprägungen in die Modellwelt einfügen lassen, gibt es dennoch in der Forschung. So könnte die Berücksichtigung von ethischen Gesichtspunkten eine sinnvolle Erweiterung der Agency-Theorie darstellen. Dabei wird davon ausgegangen, dass Agenten ein Grundlevel an Ethik aufweisen und einen negativen Nutzen erleiden, wenn sie weniger leisten, als dies mit dem Prinzipal vereinbart war (vgl. Stevens & Thevaranjan, 2010, S. 136 f.).

2.2. Verhaltenswissenschaftliche Ansätze zur Berücksichtigung von Ehrlichkeitspräferenzen

Evans et al. (2001) bauen auf dem Gedanken von Baiman und Lewis (1989) auf, dass Agenten ein Grundlevel an Ethik besitzen. Sie fanden heraus, dass Individuen entgegen der Agency-Theorie auch dann ein gewisses Level an Ehrlichkeit aufweisen, wenn ihnen keinerlei Anreiz gegeben wird, dies zu tun. Im Setting ihrer Experimente hatten sie sogar Anreize gesetzt, die die Teilnehmer zum Lügen motivieren sollten. Dabei verzichteten die Teilnehmer auf bis zu 50 Prozent des möglichen Pay-Offs (vgl. Evans et al., 2001, S. 552 f.). Ein ähnliches Ergebnis findet sich auch in einer Studie von Matuszewski (2010) wieder (vgl. Matuszewski, 2010, S. 246).

Gemäß der Prinzipal-Agenten-Theorie erleiden Agenten keinen Disnutzen, wenn sie lügen (vgl. Evans et al., 2001, S. 542). Wenn sich jedoch zeigt, dass Agenten nicht den maximal möglichen Betrag abgreifen, der ihnen laut Agency-Theorie zur Verfügung stehen würde, scheint Ehrlichkeit eine Rolle in den Überlegungen der Agenten zu spielen (vgl. Schwering, 2016, S. 88). Anders als in der Prinzipal-Agenten-Sichtweise vorherrschend, geht es Individuen offensichtlich nicht ausschließlich um die Maximierung ihres eigenen (finanziellen) Nutzens. Ein Grund dafür liegt darin, dass Individuen Fairnesspräferenzen aufweisen und damit teilweise anderen Präferenzen unterliegen, als dies die Agency-Theorie vorschlägt. Dadurch erleiden sie einen Disnutzen, wenn sie sich unehrlich verhalten, und Ehrlichkeit kann als Teil der Nutzenfunktion des Agenten angesehen werden (vgl. Luft, 1997, S. 211 f.). Eine Motivation können Individuen dementsprechend aus beiden Richtungen finden: aus der opportunistischen Agency-Betrachtung mit dem Ziel der eigenen Nutzenmaximierung als extrinsische Motivation in Form von (finanziellen) Anreizen, aber auch aus einer psychologischen Sichtweise, in der Individuen einer intrinsischen Motivation zu Werten wie Ehrlichkeit unterliegen. In der Folge wägen Individuen die beiden Faktoren gegeneinander ab. So verhalten sich auch Menschen, die sich selbst als ehrlich einschätzen, zuweilen unethisch, aber nur in dem Maße, dass sich dadurch ihre Einschätzung über ihre eigene Ehrlichkeit nicht verändert (vgl. Mazar, Amir & Ariely, 2008, S. 642). Neben Ehrlichkeit können auch weitere Motivationsquellen wie der Wunsch nach Respekt oder das Reputationsstreben von Individuen dafür sorgen, dass die Beobachtungen von den Agency-Vorhersagen abweichen (vgl. Murphy, Wynes, Hahn & Devine, 2020, S. 976).

Die Ergebnisse von Drake, Matuszewski und Miller (2014) unterstützen den Gedanken von Luft (1997), wonach Agenten eine Abwägung zwischen der persönlichen Wohlstandsmehrung und Faktoren wie Ehrlichkeit vornehmen. 68 Prozent der Studienteilnehmer nahmen diese Abwägung vor. Lediglich zwei von 85 Teilnehmern entschieden sich zu einer Maximierung ihres Wohlstandes, was gegen die rationale Nutzenmaximierer-Sicht der Agency-Theorie spricht (vgl. Drake et al., 2014, S. 65).

Zudem verweist die Studie auf 30 Prozent der Teilnehmer, die immer ehrlich berichtet hätten und die bereits genannten 2 Prozent, die immer so berichtet hätten, dass sie ihren persönlichen Wohlstand maximieren können (vgl. ebd., S. 65). Dies unterstützt einen Ansatz nach Koford und Penno (1992). Demnach gibt es ethische und eigennutzinteressierte Agenten. Während ethische Agenten immer die Wahrheit sagen, lügen eigennutzinteressierte Agenten bei jeder Gelegenheit, wenn sie dadurch ihren Wohlstand maximieren können (vgl. Koford & Penno, 1992, S. 131 f.).

Die Ergebnisse legen nahe, dass Agenten einen Trade-Off zwischen intrinsischem Nutzen aus Ehrlichkeit und extrinsischem Nutzen aus Misreporting vornehmen. Um die Funktionsweise dahinter verstehen und erklären zu können, sollte die Agency-Theorie Ehrlichkeit berücksichtigen und Prinzipale in der Gestaltung ihrer Verträge Ehrlichkeitspräferenzen entsprechend berücksichtigen (vgl. Mittendorf, 2006, S. 90). Wenn es Unternehmen gelingt die Kontextfaktoren zu verstehen, die auf die Ehrlichkeit ihrer Mitarbeiter einwirken, kann es ihnen dadurch auch gelingen selbst einen Nutzen aus dieser Präferenz zu ziehen (vgl. Douthit & Stevens, 2015, S. 488).

2.3. Budgetary Slack als Folge von Informationsasymmetrien und Aufgabendelegation

Als Budgetary Slack wird die Differenz zwischen den Planwerten aus der Budgetierung und den tatsächlich benötigten Ressourcen zur Zielerreichung bezeichnet (vgl. Kilfoyle & Richardson, 2011, S. 186). Die Differenz ergibt sich dabei durch die berichteten Budgets der Manager im Unternehmen. Das Anreizsystem eines Managers ist häufig an dessen Performance im Vergleich zu den Budgetwerten gekoppelt. Schneidet er besser als ein bestimmter Zielwert ab, so erhält er eine (finanzielle) Belohnung. Manager haben deshalb einen Anreiz, zu versuchen ihre Budgets so festzulegen, dass die Ziele möglichst leicht zu erreichen sind. Im Falle eines Kostenbudgets könnte ein Manager bspw. höhere Kosten melden, als er tatsächlich erwartet (vgl. Paul, 2014, S. 20). Für den Manager bietet sich also die Möglichkeit von Slack persönlich zu profitieren (vgl. Church, Kuang & Liu, 2019, S. 74). Die Bildung von Slack durch den Manager erfolgt dabei absichtlich (vgl. Dunk & Perera, 1997, S. 649). Abb. 1 verdeutlicht die Slack-Bildung sowohl für den Fall von Kosten- als auch Leistungsbudgets. Im Falle eines Kostenbudgets könnte ein Manager z. B. Kosten in Höhe von zehn Einheiten übermitteln, obwohl er nur mit Kosten in

Höhe von acht Einheiten rechnet. Wird ihm das Budget von der Unternehmenszentrale bewilligt, so könnte er dadurch zwei Einheiten Slack als Puffer für sich vereinnahmen.

Die Bildung von Slack kann dabei zu einer Reihe von negativen Effekten für das Unternehmen führen (vgl. Webb, 2002, S. 361). Gerade in großen dezentralisierten Unternehmen mit einem hohen Grad an Informationsasymmetrie ist die Gefahr von Moral Hazard-Problemen besonders groß (vgl. Abernethy, Bouwens & Van Lent, 2004, S. 561). Je nach Anreizgestaltung im Unternehmen können Budgets sehr eng mit der Beurteilung des verantwortlichen Managers zusammenhängen. So kann zumindest ein Teil der Bezahlung des Managers, aber auch Faktoren wie die zukünftige Ressourcenzuteilung und Beförderungschancen, von der Performance des Managers relativ zu gemeldeten Budgetwerten abhängen. Dies schafft für den Manager einen Anreiz für sich möglichst viel Slack zu verhandeln (vgl. Van der Stede, 2000, S. 609). Dadurch entsteht eine Verzerrung in der Performance-Beurteilung des Managers. Auch ein verringerter Arbeitsaufwand von Seiten des Managers in Folge des als Puffer gebildeten Slacks ist denkbar. Zudem können verzerrte Berichte an die Unternehmenszentrale zu einer Fehlallokation von Ressourcen führen (vgl. Webb, 2002, S. 361).

Gleichzeitig gibt es die Sichtweise, dass Slack unter bestimmten Bedingungen auch für das Unternehmen vorteilhaft sein kann. So könnten Unternehmen die Bildung von Slack bewusst zulassen, um den Manager in einer Situation mit mehreren möglichen Zielen zur Verfolgung des aus Unternehmenssicht wichtigsten Ziels zu motivieren. Insbesondere wenn mehrere kurzfristige Ziele gleichzeitig vorliegen, könnte Slack somit für das Unternehmen die Möglichkeit bieten, das Verhalten des Managers zu steuern (vgl. Davila & Wouters, 2005, S. 606 f.). Außerdem kann Slack in solchen Situationen den Aufwand des Agenten steigern und sich damit positiv auf die Unternehmensperformance auswirken (vgl. Schatzberg & Stevens, 2008, S. 78).

In der Praxis ist die Neigung zur Slack-Bildung von einer Vielzahl von Faktoren abhängig. Werden bspw. mehrperiodische Budgetsituationen untersucht, zeigt sich, dass der von Managern berichtete Slack geringer ist, als er gemäß der Agency-Theorie sein sollte. Einen Erklärungsansatz dafür bietet die Betrachtung von Reputationsbeziehungen innerhalb des Unternehmens. Die Reputation eines Managers mit Budgetverantwortung wird sich u. a. danach orientieren, wie gut seine Ist-Werte im Vergleich zu den berichteten Soll-Werten abschneiden. Schneidet ein Manager besser als sein berichtetes Budget ab, so wird er eine bessere Reputation aufbauen. Gleichzeitig kann hier auch ein umgekehrter Effekt beobachtet werden. Wenn ein Manager über einen längeren Zeitraum wesentlich besser als die berichteten Budgets abschneidet, so wird die Zentrale die Berichte des Managers als unzuverlässig einschätzen und der Manager in der Folge einen Reputationsverlust erleiden (vgl. Webb, 2002, S. 363 f.). Für Manager kann der Aufbau einer positiven Reputation langfristig größere Vorteile mit sich bringen (z. B. durch eine Beförderung in der Zukunft), als das kurzfristige Ausnutzen von finanziellen Vorteilen in Form von Slack (vgl. ebd., S. 375).

Wenn Manager unter dem Vorliegen von Informationsasymmetrien im Unternehmen ihre privaten Informationen an den Prinzipal weitergeben, unterliegen sie folglich einem Spannungsfeld aus den erzielbaren Vorteilen aus der wahrheitsgemäßen Weitergabe ihrer privaten Information und den erzielbaren Vorteilen aus dem Zurückhalten oder Falschberichten von Informationen (vgl. Hannan et al., 2006, S. 889). Eine ehrliche Berichterstattung des Agenten entspricht somit grundsätzlich dem Gegenteil einer absichtlichen Slack-Bildung. Eine Möglichkeit Agenten zu dieser ehrlichen Art der Berichterstattung zu motivieren, stellen bspw. wahrheitsinduzierende Vergütungssysteme dar. Aufgrund von Informationsasymmetrien und unterschiedlichen Zielen lässt sich eine vollständig wahrheitsgemäße Berichterstattung in der Realität jedoch nur schwer verwirklichen (vgl. Daumoser, Hirsch & Sohn, 2018, S.118). Dennoch scheint es für ein Unternehmen aufgrund der potenziellen negativen Folgen durch die Bildung von Slack wichtig zu sein, zu verstehen welche Faktoren helfen, um die Slack-Bildung im Unternehmen zu verringern (vgl. Sprinkle & Williamson, 2006, S. 419).

2.4. Das object-of-control-Framework nach Merchant/Van der Stede

Im Folgenden soll das auf Merchant und Van der Stede (2017) zurückgehende object-of-control-Framework näher erläutert werden. Merchant und Van der Stede (2017, S. 12-14) benennen drei Gründe, die das Auftreten von Management Control-Problemen im Unternehmen erklären: Lack of Direction, Motivational Problems und Personal Limitations. Lack of Direction bezeichnet das Problem, dass Mitarbeiter nicht wissen, was das Unternehmen von ihnen erwartet. Motivational Problems treten auf, wenn die Ziele des Mitarbeiters von den Zielen des Unternehmens abweichen. Ursachen hierfür finden sich in der Eigennutzorientierung der Mitarbeiter. Anders verhält es sich bei Personal Limitations. Hier wissen die Mitarbeiter, was von ihnen erwartet wird und sind motiviert im Sinne des Unternehmens zu handeln, jedoch verhindern persönliche Einschränkungen, dass die Ziele tatsächlich erreicht werden. Gründe dafür können bspw. fehlendes Wissen oder eine fehlende Ausbildung des Mitarbeiters sein. Dies deckt sich mit Überlegungen aus der Prinzipal-Agenten-Theorie. Die Möglichkeit zu opportunistischen Verhalten wird dem Manager hier durch die Übertragung von Entscheidungsrechten ermöglicht. Die Übertragung der Entscheidungsrechte resultiert aus dem Auftreten von Informationsasymmetrien im Unternehmen. Je stärker Informationsasymmetrien ausgeprägt sind, desto größer ist auch die Gefahr, dass Agenten ihre Entscheidungsrechte zu eigenen Zwecken missbrauchen und desto größer ist in der Folge der Bedarf des Prinzipals diesem Verhalten entgegenzuwirken (vgl. Fiolleau, Libby & Thorne, 2018, S. 120).

Treten im Unternehmen Steuerungsprobleme auf, sollte zunächst überprüft werden, ob es möglich ist das Auftreten der Probleme zu vermeiden. Mögliche Strategien für das Unternehmen liegen hier z. B. in der Eliminierung bestimmter Tätigkeiten durch ein Auslagern an Drittparteien oder einer

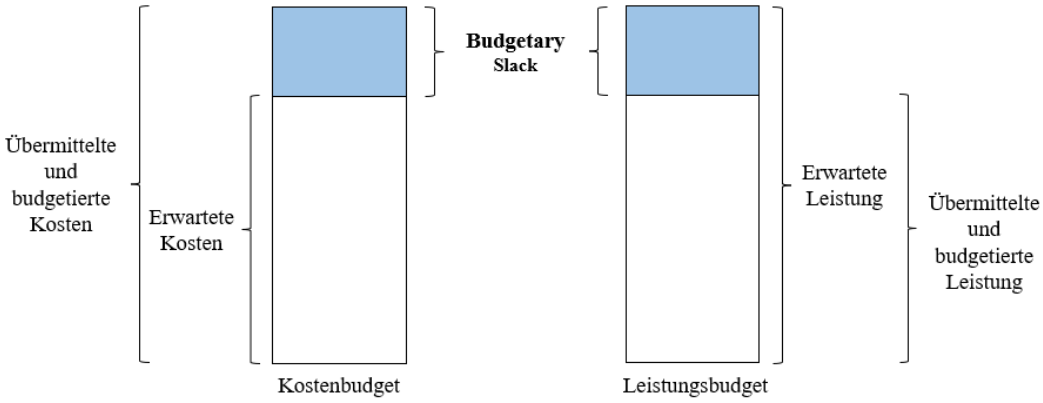

Abbildung 1: Budgetary Slack
Quelle: In Anlehnung an Schwering, 2016, S. 30

Automatisierung bestimmter Schritte (vgl. Merchant & Van der Stede, 2017, S. 15 f.).

Wenn ein Vermeiden der Control-Probleme nicht möglich ist, sollte ein Unternehmen auf Steuerungsinstrumente zurückgreifen, wobei in der Gesamtheit aller Instrumente im Unternehmen vom MCS gesprochen wird (vgl. Merchant & Van der Stede, 2017, S. 19). Das Unternehmen kann hierbei nach Merchant und Van der Stede (2017) auf eine Reihe verschiedener Controls zurückgreifen, die sich in vier übergreifende Control-Arten unterteilen. Diese betreffen die erzielten Ergebnisse der Mitarbeiter (Results Controls), die von ihnen durchgeführten Handlungen (Action Controls) oder beziehen sich auf die Einstellung und Ausbildung von Mitarbeitern (Personnel Controls) bzw. im Unternehmen vorherrschende Wertvorstellungen (Cultural Controls) (vgl. Merchant & Van der Stede, 2017, S. 19). Results Controls und Action Controls können den formellen Steuerungsinstrumenten zugeordnet werden, während Personnel Controls und Cultural Controls informelle Steuerungsinstrumente darstellen (vgl. Feichter & Grabner, 2020, S. 149). Informelle Controls beschreiben eine eher unterbewusste Art der Verhaltenssteuerung. Sie sind nur schwer zu implementieren und eine Verhaltensänderung der Mitarbeiter durch informelle Controls benötigt Zeit, gleichzeitig können sie durch ihre Existenz auch einen Rahmen schaffen, der die Anwendung formeller Steuerungsinstrumente positiv verstärkt (vgl. Malmi & Brown, 2008, S. 295).

Manager können der Anwendung von Control-Instrumenten im Unternehmen negativ gegenüberstehen. Denkbar ist, dass diese ihr Verhalten in dem Sinne anpassen, dass Controls vermieden werden. Agenten empfinden das Vorhandensein von Steuerungsinstrumenten als ein Zeichen des Misstrauens von Seiten des Prinzipals und fühlen sich dadurch in ihren Entscheidungsmöglichkeiten eingeschränkt (vgl. Falk & Kosfeld, 2006, S. 1628). Dadurch kann die Nutzung von Kontrollmechanismen opportunistisches Verhalten auslösen, während andererseits fehlende Kontrollmechanismen vorherrschende (negative) Werte und Normen der Manager weiter verstärken können (vgl. Tayler & Bloomfield, 2011, S. 783). Wichtig ist folglich eine Abwägung zwischen dem Nutzen von Controls und den Hidden Costs, die bei Anwendung der Controls auftreten. Dabei sollte das Unternehmen beachten, dass Agenten sehr heterogene Charaktereigenschaften aufweisen und Steuerungsinstrumente gerade bei sehr opportunistisch veranlagten Agenten unabdingbar sind. Außerdem sollte das Unternehmen bei der Gestaltung seines MCS versuchen kein negatives Verhalten zu begünstigen (vgl. Falk & Kosfeld, 2006, S. 1628). Der Kontingenztheorie folgend sind Individuen nur beschränkt rational. Entsprechend schwierig ist es für Unternehmen Anreize zu setzen, um das Verhalten der Mitarbeiter auf die Unternehmensziele abzustimmen. So könnten Unternehmen bei der Gestaltung ihres MCS Fehler unterlaufen und die Reaktion der Manager darauf ist nicht vorhersehbar. Abweichungen zwischen den tatsächlichen Aktionen des Agenten zu den vom Unternehmen erwünschten Aktionen könnten demnach unbeabsichtigt auftreten und auf Entscheidungsfehler zurückzuführen sein (vgl. Covaleski et al., 2003, S. 30).

Trotz der mit der Implementierung eines MCS verbundenen Kosten tragen effektiv ausgestaltete Steuerungsinstrumente dazu bei Steuerungsprobleme zu lösen und verringern opportunistisches Verhalten der Mitarbeiter. Ein Unternehmen steht dabei bei der Implementierung von Steuerungsinstrumenten vor einer Kosten-Nutzen Abwägung (vgl. Feichter & Grabner, 2020, S. 154). Möchten sich Unternehmen die grundsätzliche Tendenz zur Ehrlichkeit von Individuen zunutze machen, sollten sie evaluieren, wie sich verschiedene Arten von Beziehungen zwischen ihnen und den Mitarbeitern auf das Verhalten der Mitarbeiter auswirken. Unternehmen sollten hierbei bedenken, dass sich die Tendenz zur Ehrlichkeit durch eine Vielzahl von Kontextfaktoren verstärken, aber auch abschwächen lässt (vgl. Evans et al., 2001, S. 553).

3. Formelle Steuerungsinstrumente zur Sicherstellung einer wahrheitsgemäßen Berichterstattung

3.1. Effektivität von Results Controls

Zu der Gruppe der Results Controls werden jene Steuerungsmechanismen gezählt, die Mitarbeiter für das Erreichen gewünschter Ziele belohnen (Merchant & Van der Stede, 2017, S. 33). Klar definierte Ziele im Rahmen von Results Controls sorgen für ein Verständnis, welche Ziele aus Unternehmenssicht wünschenswert sind und motivieren Mitarbeiter zur Erreichung der Ziele beizutragen. Dadurch können Results Controls einen Beitrag zur Beseitigung von Lack of Direction leisten. Durch die Verknüpfung der gewünschten Ziele mit Anreizen für den Mitarbeiter werden auch Motivational Problems adressiert (vgl. Merchant & Van der Stede, 2017, S. 37).

Das bloße Vorhandensein von Informationsasymmetrien allein hat grundsätzlich noch keinen Einfluss darauf, ob ein Unternehmen auf ein Anreizsystem im Rahmen von Results Controls zurückgreifen sollte. Vielmehr geht die Anwendung dieser mit dem Grad der Delegation von Entscheidungsrechten einher. Dabei könnte die Delegation von Entscheidungsrechten jedoch auch als Reaktion des Unternehmens auf die Informations-asymmetrien angesehen werden, wodurch der Grad an Informationsasymmetrie zwar keinen direkten, aber einen indirekten Einfluss auf die Verwendung von Results Controls ausübt (vgl. Abernethy et al., 2004, S. 562 f.). Die Delegation von Entscheidungsrechten kann zu einer Vielzahl positiver Faktoren beitragen, wie bspw. reduzierten Kosten der Informationsgewinnung oder einer besseren Unternehmensperformance. Doch gleichzeitig ergibt sich durch die Übertragung von Entscheidungsrechten die Möglichkeit für den Manager, diese Situation auszunutzen und sich durch unethisches Verhalten einen Nutzen zu erarbeiten, was dem Unternehmen schadet. Aus Agency-Sicht stellt genau diese Situation ein Problem für das Unternehmen dar, da der Manager aus Selbstnutzenorientierung die Situation für sich ausnutzen wird, wenn das Unternehmen nicht entsprechende Anreize schafft, sich ethisch zu verhalten (vgl. Chong & Wang, 2019, S. 275). Gerade wenn der Manager über eine Vielzahl von Entscheidungsbefugnissen verfügt, erscheint die Anreizgestaltung in Form von Results Controls als ein sinnvolles Instrument, um sein Verhalten im Sinne des Unternehmens zu steuern. Die Unternehmensführung stützt sich bei der Ausgestaltung ihres Anreizsystems auf das beobachtete Reporting-Verhalten des Managers in der Vergangenheit und wird aufgrund der motivierenden Wirkung ein variables Anreizsystem, bei dem das vom Manager erzielbare Gehalt an das Erreichen von für das Unternehmen wünschenswerten Zielen gekoppelt ist, einem Festgehalt vorziehen (vgl. Cardinaels & Yin, 2015, S. 1011). Die vorherrschende Heterogenität unter den Managern im Unternehmen macht es zu einer Herausforderung für Unternehmen, effektive Anreizsysteme zu implementieren. Auf der einen Seite sollte das Verhalten von zu Opportunismus neigenden Mitarbeitern mit Anreiz- und Belohnungssystemen gesteuert werden, auf der anderen Seite sollten diese Maßnahmen nicht zu einer demotivierenden Wirkung auf die Mitarbeiter führen, die zu ehrlichem Verhalten neigen (vgl. Nagin et al., 2002, S. 870).

Unternehmen neigen zu einer dezentralen Steuerung, wenn der Grad an Informations-asymmetrie im Unternehmen hoch ist. Um die Leistung ihrer Mitarbeiter zu bewerten und zu entlohnen, greifen Unternehmen in solchen Situationen oft auf verschiedene Kennzahlen und Ziele zurück (vgl. Abernethy et al., 2004, S. 547). Um Results Controls effektiv zu gestalten, muss dem Manager bekannt sein, welche Ziele das Unternehmen verfolgt. Dazu sollte sich das Unternehmen bewusst machen, welche Ziele Priorität haben und dies klar an die Mitarbeiter kommunizieren (vgl. Merchant & Van der Stede, 2017, S. 42). Unter unsicheren Umweltbedingungen sind die Kosten der Informationsgewinnung für das Unternehmen hoch. Organisationseinheiten näher am operativen Geschäft haben mehr Informationen, Unternehmen können diese Informationen effizient nutzen und gleichzeitig hohe Kosten der Informationsgewinnung vermeiden. Durch die Kommunikation von gewünschten Zielen wird versucht einen Zielkongruenz zwischen Unternehmen und Manager zu erreichen, wodurch die Gefahr opportunistischen Verhaltens und damit verbundenen höheren Kosten für das Unternehmen, minimiert werden soll (vgl. Nagar, 2002, S. 381). In der Prinzipal-Agenten-Theorie wird dem Agenten grundsätzlich unterstellt, dass er risiko- und arbeitsscheu sei. Ein Anreizsystem stellt hier eine Motivationskomponente dar, um das Verhalten des Agenten in die gewünschte Richtung zu lenken (vgl. Chong & Wang, 2019, S. 279 f.). Results Controls können den Agenten für übernommenes Risiko entlohnen und verstärken dadurch die Zielkongruenz. So zeigt sich, dass die Tendenz von risikoaversen Agenten in Richtung einer höheren Slack-Bildung weist. Slack kann demnach als eine Antwort auf Ungewissheit charakterisiert werden (vgl. Young, 1985, S. 840). Da ein zu großer Druck die Tendenz zum Lügen verstärkt, sollten Unternehmen überprüfen, ob die festgelegten Ziele realistisch zu erreichen sind (vgl. Grover, 2005, S. 156). Einen Ansatzpunkt für das Unternehmen kann hier die SMART-Formel bieten, wonach Ziele spezifisch, messbar, akzeptiert, realistisch und terminiert sein sollten (vgl. Doran, 1981, S. 36).

Je volatiler, größer und innovativer Unternehmen sind, desto mehr Entscheidungsrechte geben diese an ihre Manager ab. Gleichzeitig werden Manager mit höheren Entscheidungsbefugnissen in der Regel auch mit einem höheren Anteil von variablen Bonus-Bestandteilen vergütet (vgl. Nagar, 2002, S. 391). Durch die Übertragung von Entscheidungsrechten auf den Agenten wird dessen Aktionsraum erweitert, gleichzeitig steigt aber auch dessen Verantwortung für die von ihm vorgenommenen Aktionen an (vgl. Chong & Wang, 2019, S. 283). Klassischerweise handelt es sich bei Mitarbeitern mit Budgetverantwortung um höherrangige Manager im Unternehmen. Das Prinzip der Controllability scheint damit für solche Mitarbeiter gut erfüllbar zu sein. Controllability wird von Merchant und Van der Stede (2017) als eine der Voraussetzungen genannt, um Results Controls effektiv gestalten zu können. Demnach sollten Mitarbeiter in der Lage sein, die Ergebnisse, nach denen sie beurteilt werden, auch

aktiv zu beeinflussen (vgl. Merchant & Van der Stede, 2017, S. 42 f.). Werden Budgets zur Verteilung knapper Ressourcen im Unternehmen genutzt, ergibt sich für den Manager ein Anreiz seinen tatsächlichen Ressourcenbedarf zu überschätzen (vgl. Fisher, Maines, Peffer & Sprinkle, 2002, S. 848). Wird das Budget zusätzlich für die Anreizgestaltung genutzt, lässt sich dadurch die Slack-Bildung reduzieren. Zwar berichten die Manager ein zu hohes Budget, dies ist aber mit einer erhöhten Performance verbunden. Der Grund dafür lässt sich zum einen in der gesteigerten Motivation der Manager finden, mehr von den knappen Ressourcen zu erhalten, und zum anderen auf die intrinsische Motivation dieser zurückführen, die selbst abgegebenen Vorgaben zu übertreffen (vgl. Fisher et al., 2002, S. 862). Wenn sich Unternehmen darauf konzentrieren, dass ihr MCS die Möglichkeiten zur Slack-Bildung eingrenzt, so kann dies andere für das Unternehmen ungünstige Faktoren verstärken. Eine denkbare Reaktion von Managern, die sich weniger Slack in ihren Verhandlungen erarbeiten können, wäre es ihren Fokus stärker auf kurzfristigen Unternehmenserfolg zu legen, um die gesetzten Ziele erreichen zu können (vgl. Van der Stede, 2000, S. 619). Unternehmen sollten entsprechend beachten wie sich die gesetzten Ziele auf das Verhalten der Mitarbeiter auswirken und welche weiteren Auswirkungen dies auf weiterführende Entscheidungen des Managements ausübt. Ob ein anreizbasiertes Vergütungssystem in diesem Kontext bei ohnehin unehrlichen Managern ein wirksames Mittel ist, um Misreporting zu vermeiden, wird von Chong und Wang (2019) angezweifelt. Hier erscheint es für die beiden Autoren ein wirksameres Mittel zu sein, die Entscheidungsrechte der Manager zu beschränken, so dass für die Manager die Möglichkeiten zu unethischem Verhalten minimiert werden (vgl. Chong & Wang, 2019, S. 299).

Oftmals wird die Performance eines Unternehmensbereichs durch Größen gemessen, die den ganzen Bereich abbilden, wie z. B. den Bereichsgewinn oder den Return on Investment, den dieser Bereich erwirtschaftet hat. Dabei ergibt sich jedoch unter Umständen eine Problematik dadurch, dass sich die Entscheidungen mehrerer Manager gleichzeitig auf die gewählte Entscheidungsgröße auswirken (vgl. Abernethy et al., 2004, S. 546). Sind die Ziele im Unternehmen klar definiert und die Manager in der Lage zur Erreichung der Ziele beizutragen, nennen Merchant und Van der Stede (2017) als letzte Notwendigkeit zur effektiven Ausgestaltung von Results Controls die Eigenschaft, dass die Ergebnisse auch effektiv gemessen werden können (vgl. Merchant & Van der Stede, 2017, S. 43). Dies kann insbesondere dann ein Problem darstellen, wenn im Unternehmen starke Leistungsverflechtungen vorliegen. Dabei kann die betrachtete Größe von anderen Bereichen beeinflusst werden und in der Folge nur ein stark verzerrtes Bild der tatsächlichen Performance des mit der Größe beurteilten Managers wiedergeben. Wirkt sich der Einfluss anderer Bereiche negativ auf die Entlohnung des Managers aus, so demotiviert das diesen. In der Folge kann eine enge Leistungsverflechtung somit möglicherweise dazu führen, dass Manager für das Unternehmen notwendige Kooperationen unterlassen und mögliche Synergien ausbleiben (vgl. Abernethy et al., 2004, S. 549). Dabei kann zusätzlich danach unterschieden werden, ob der betrachtete Bereich beeinflusst wird oder ob dieser auf einen anderen Bereich Einfluss nimmt. Wenn der betrachtete Bereich einen anderen Bereich beeinflusst, sind Kennzahlen auf Bereichsebene weniger dazu geeignet die Performance des Managers abzubilden. Hier sollte auf übergreifende Größen zurückgegriffen werden (z. B. der Gewinn der gesamten Business Unit). Anders verhält es sich, wenn der betrachtete Bereich einen Einfluss durch andere Bereiche erfährt. Hier kann die Verwendung von Kennzahlen auf Bereichsebene anzeigen, wie gut ein Manager mit Leistungsverflechtungen umgehen kann (z. B. die Fähigkeit zum Heben von Synergien) (vgl. Abernethy et al., 2004, S. 562). Merchant und Van der Stede (2017) beschreiben deshalb neben dem Prinzip der Controllability und einer Zielkongruenz die folgenden Eigenschaften zur effektiven Gestaltung von Results Controls: Die Performance-Messung sollte präzise, objektiv, zeitbezogen, verständlich und kosteneffizient erfolgen. Unternehmen sollten bei der Gestaltung von Results Controls bedenken, dass diese unethisches Verhalten auch begünstigen können. Je höher der finanzielle Anreiz, der Managern gesetzt wird, desto wahrscheinlicher ist es auch, dass diese nicht wahrheitsgemäß berichten. Ist der zu erwartende Bonus wiederum sehr gering, tendieren Manager eher nicht zur Lüge. Der Bonus kann also als eine Art der Kompensation für das Lügen angesehen werden (vgl. Cardinaels & Jia, 2016, S. 200).

In der Praxis zeigt sich, dass die Möglichkeit Slack auszubilden, immer auch stark von der voran gegangenen Performance abhängt. So sind Abteilungen, die sich in der Vergangenheit als sehr profitabel herausgestellt haben, in der Regel weniger starken Control-Instrumenten unterstellt, und haben damit auch bessere Möglichkeiten Slack zu bilden (vgl. Van der Stede, 2000, S. 619). Stevens (2002) versucht in seiner Studie herauszufinden, warum sich selbst bei Verträgen, die den Mitarbeiter zur Bildung von Slack motivieren, weniger Slack finden lässt, als dies nach Agency-Theorie vorhergesagt wäre. Dazu untersucht er den Einfluss von Reputationseffekten und Ethik auf die Slack-Bildung. Dabei findet er heraus, dass sich beide Faktoren negativ auf die Bildung von Slack auswirken. So haben Manager mit einem hohen Reputationsgedanken eine stärkere Wahrnehmung davon, dass Slack unethisch ist (vgl. Stevens, 2002, S. 169). Brown, Evans und Moser (2009) untersuchen eine Reihe von Studien danach, ob das Ergebnis der Studien konsistent mit den nach der Agency-Theorie erwarteten Ergebnissen ist. Dabei entsprechen viele Studien auch dem erwarteten Ergebnis aus der Agency-Theorie, gleichzeitig finden die Autoren aber auch eine Reihe von Hinweisen auf Anomalien. So finden sie z. B. in den Studien weniger Slack vor, als dies die Agency-Theorie vorhersagen würde oder entdecken, dass sozialer Druck einen Einfluss auf die Slack-Bildung nimmt. Diese Anomalien sollten folglich weiter untersucht werden, um diese gegebenenfalls in das Agency-Modell miteinfließen lassen zu können (vgl. Brown et al., 2009, S. 339-342). Auch Chong und Wang (2019) sind geteilter Meinung, was die An-

wendbarkeit der Agency-Theorie in der Forschung betrifft. Sie verweisen dazu darauf, dass Studien gezeigt hätten, dass die Gestaltung von Anreizsystemen sogar zu einem höheren Level an Misreporting führt (vgl. Chong & Wang, 2019, S. 276).

3.2. Anreizsysteme im Rahmen von Results Controls

Die Implementierung von Results Controls erfolgt gemäß Merchant und Van der Stede (2017, S. 38) in vier Schritten. Ausgehend von der Definition der Performance-Dimensionen wird die Performance gemessen. Zudem werden den Mitarbeitern Performance-Ziele gesetzt und abschließend Mitarbeiter für die Erreichung der Performance-Ziele belohnt. Auf diesen vier Elementen aufbauend wird im Folgenden untersucht, wie im Rahmen der Implementierung von Results Controls zu einer wahrheitsgemäßen Berichterstattung im Unternehmen beigetragen werden kann.

Definition der Performance-Dimensionen

Effektive Results Controls setzen eine Zielkongruenz zwischen Prinzipal und Agent voraus. Das langfristige Ziel eines Unternehmens ist es, den Unternehmenswert zu maximieren (vgl. Feichter & Grabner, 2020, S. 155). Im Zuge dessen wird häufig ein Shareholder-Ansatz verfolgt, bei dem die Ziele des Unternehmens und damit auch das MCS des Unternehmens darauf ausgerichtet sind, die Rendite der Shareholder zu erhöhen (vgl. Ittner & Larcker, 2001, S. 368). Gerade in höheren Hierarchieebenen eignen sich deshalb Budgets als Performance-Dimension. Das Budget als das Ergebnis der Planung ist darauf ausgerichtet den Wert des Unternehmens zu steigern (vgl. Ewert & Wagenhofer, 2014, S. 400). Allerdings bietet die Nutzung von Budgets Managern auch einen Anreiz zur Falschberichterstattung. Misreporting kann dabei bspw. in Form von Überschätzen der erwarteten Kosten oder auch dem Unterschätzen der erwarteten Performance erfolgen (vgl. Schwering, 2017, S. 1152). In niedrigeren Ebenen sollte das Unternehmen darauf achten, dass durch die Definition der Performance-Dimension das Controllability-Prinzip nicht verletzt wird. Hier sind Dimensionen wie etwa die Planungsgenauigkeit (vgl. Fisher et al., 2002, S. 862) oder die Weitergabe von Wissen im Unternehmen (vgl. Haesebrouck, Cools & Van den Abbeele, 2018, S. 230) denkbar.

Performance-Messung

Im Rahmen der Performance-Messung geht es im Unternehmen zunächst darum, die Ziele des Unternehmens in messbare Kennzahlen zu übersetzen. Die Kennzahlen sollten dazu geeignet sein, zu überprüfen, ob die von den Mitarbeitern durchgeführten Aktionen dazu beigetragen haben, die Ziele des Unternehmens zu erreichen (vgl. Feichter & Grabner, 2020, S. 156). Kennzahlen sollten dabei dieselben Anforderungen erfüllen, die für Results Controls allgemein vorausgesetzt werden (vgl. Merchant & Van der Stede, 2017, S. 39). In der Praxis werden Kennzahlen aus den folgenden Performance-Bereichen am meisten eingesetzt (in absteigender Reihenfolge): Messgrößen auf Bereichsebene wie etwa der Bereichsumsatz, Messgrößen auf Unternehmensebene wie der Unternehmensumsatz und der Aktienkurs (vgl. Keating, 1997, S. 267). Um die Kooperation zwischen verschiedenen Unternehmensbereichen und -managern zu verbessern, eignen sich Kennzahlen auf Gruppenebene besser als Kennzahlen, die lediglich auf die individuelle Performance abstellen (vgl. Jensen, 2003, S. 396). Kennzahlen können auf finanziellen und nichtfinanziellen Größen aufsetzen. Einer ökonomischen Sichtweise folgend sollte es keinen Einfluss auf das Verhalten des Managers haben, ob er nach finanziellen oder nichtfinanziellen Kennzahlen bewertet wird. Aus einer psychologischen Perspektive heraus kann diese Unterscheidung jedoch die wahrheitsgemäße Berichterstattung von Managern beeinflussen. So nimmt die Ehrlichkeit von Menschen ab, wenn sie nach finanziellen Kennzahlen beurteilt werden (vgl. Church et al., 2019, S. 82).

Eine Möglichkeit stellt es dar, das Anreizsystem auf dem vom Agenten gemeldeten Budget aufzusetzen. Im Zuge der Zielvereinbarung werden anhand einer Kennzahl (z. B. Umsatz) Regeln aufgestellt, wie der Agent belohnt (bestraft) wird, wenn er das Budgetziel übertrifft (nicht erreicht) (vgl. Covaleski et al., 2003, S. 14). Im Rahmen der partizipativen Budgetierung erhält das Unternehmen vom Agenten private Informationen. Dem Unternehmen ist dabei bewusst, dass der Agent einen Informationsvorsprung besitzt und deshalb die Gefahr der Slack-Bildung besteht, weshalb das Unternehmen vor der Entscheidung steht, die Ergebnisse aus der Budgetierung für die Anreizgestaltung zu verwenden oder nicht (vgl. ebd., S. 15).

Setzen von Zielvorgaben

Wird das Budget mit dem Anreizsystem eines Managers verknüpft, so führt dies in der Regel zu zwei Effekten. Zum einen wird der Manager versuchen ein möglichst leicht zu erreichendes Ziel zu verhandeln, da es dadurch für ihn leichter wird einen Bonus zu erzielen. Zum anderen wird der Manager alles dafür tun, um die festgelegte Zielgröße zu erreichen, selbst wenn er dadurch dem ganzen Unternehmen schadet (vgl. Jensen, 2003, S. 382). Ein einfaches Beispiel hierzu wäre das Vorziehen von Umsätzen aus dem Folgejahr, um die Betrachtungsperiode für den Bonus positiv zu beeinflussen.

Für die Festlegung eines Zielwertes spielen das angestrebte Zielniveau sowie die Informationen, die in die Zielfindung miteinfließen, eine zentrale Rolle (vgl. Feichter & Grabner, 2020, S. 158). Um den Einfluss wahrheitsgemäßer Berichterstattung zu untersuchen, stellt der Prozess der Zielfindung einen elementaren Ansatzpunkt dar, da durch die Berichte der Manager sowohl die Wohlfahrt des Agenten als auch die des Prinzipals beeinflusst werden können (vgl. Brown et al., 2009, S. 338 f.).

Zur Untersuchung von Ehrlichkeit und der Bildung von Slack wird häufig zwischen einem Slack-induzierenden Vertrag und einem wahrheitsinduzierenden Vertrag unterschieden. Bei einem Slack-induzierenden Vertrag wird der Manager mit einem Fixgehalt und einer Bonuskomponente vergütet. Die Höhe der Bonuskomponente ist davon abhängig, wie gut der Manager im Vergleich zu einem definierten Zielniveau abschneidet. Aufgrund vorherrschender Informationsasymmetrien hat der Manager in diesem Fall einen Anreiz zur Slack-Bildung (vgl. Libby, 2015, S. 146; Liessem, Schedlinsky, Schwering & Sommer, 2015, S. 82). Vertraut

der Prinzipal den Berichten des Managers bei Anwendung eines Slack-induzierenden Vertrages kann auch von einem Trust-Kontrakt gesprochen werden. Bspw. könnte im Kontext der Budgetierung der Manager ein Budget einreichen und der Prinzipal dieses ohne weitere Prüfung akzeptieren, was auch beiden Parteien so bewusst ist. Dies gibt dem Agenten die letzte Entscheidung über die Zahlen, die er an den Prinzipal meldet. Der Agency-Theorie folgend müsste ein Manager unter einem Trust-Kontrakt seine Berichte so ausgestalten, dass er das Maximum an Slack beanspruchen kann. In der Forschung ist die Verwendung von Trust-Kontrakten weit verbreitet, da sich dadurch Unterschiede im Verhalten der Manager zur Agency-Theorie sehr gut untersuchen lassen (vgl. Brink, Coats & Rankin, 2018, S. 93). Ein wahrheitsinduzierender Vertrag ist ein Vertrag, der durch Belohnungen und Bestrafungen auf Zielabweichungen so ausgestaltet wird, dass für den Manager die wahrheitsgemäße Weitergabe von Informationen im Rahmen eines solchen Vertrages die beste Option darstellen soll (vgl. Waller, 1988, S. 88). Ein Beispiel für einen wahrheitsinduzierenden Vertrag stellt das Weitzman (1976) dar. Wahrheitsinduzierende Anreizschemata finden in der Praxis eher selten Anwendung, da die Implementierung dieser teuer ist. Stattdessen greifen Unternehmen oft auf Verträge mit einer Bonus-Komponente zurück (vgl. Libby, 2015, S. 161).

Läge keine Informationsasymmetrie vor, würde die Wahl der Vertragsart für die Bildung von Slack keine Rolle spielen. So haben Chow, Cooper und Waller (1988) die Slack-Generierung unter verschiedenen Arten von Verträgen untersucht. Sie vergleichen einen wahrheitsinduzierenden Vertrag mit einem Slack-induzierenden Vertrag. Bei Informationssymmetrie unterscheidet sich der von den Managern berichtete Slack bei beiden Vertragsarten nicht, bei Vorhandensein von Informationsasymmetrie zeigt sich, dass die Berichte von Managern unter einem wahrheitsinduzierenden Vertrag wesentlich weniger Slack enthalten (vgl. Chow et al., 1988, S. 119 f.). Auch Chow, Cooper und Haddad (1991) sind der Meinung, dass bei wahrheitsinduzierenden Anreizsystemen grundsätzlich am wenigsten Slack gebildet wird. Bei einer mehrperiodigen Betrachtung stellt sich ein wahrheitsinduzierender Effekt aber allein dadurch ein, dass dem Prinzipal die Aktionen des Managers aus den Vorperioden bekannt werden. Die Autoren der Studie begründen dies damit, dass der Prinzipal die Slack-Bildung von Seiten des Agenten in seinen Entscheidungen antizipiert, während der Agent keine präventiven Erhöhungen des von ihm berichteten Slacks vornimmt. Als Control-Alternative zu wahrheitsinduzierenden Verträgen könnte der Prinzipal folglich auch auf einen Vertrag mit einer Ratchet (engl. für Sperrklinge) zurückgreifen, um die Slack-Bildung zu reduzieren (vgl. Chow et al., 1991, S. 57 f.). Bei Nutzung einer Ratchet werden Zielwerte von der Zielerreichung in Vorperioden abhängig gemacht. Ein hohes Zielniveau kann demnach das Ergebnis eines guten Ergebnisses in der Vorperiode darstellen. Dabei stehen die Ergebnisse von Chow et al. (1991) konträr zur herrschenden Meinung, dass sich der Ratchet-Effekt eher negativ auf die wahrheitsgemäße Berichterstattung auswirkt (vgl. Ewert & Wagenhofer, 2014, S. 466).

Eine Studie von Hobson, Mellon und Stevens (2011) untersucht verschiedene Anreizsysteme danach, ob sich die Gestaltung des Anreizsystems auf die moralische Beurteilung von Slack auswirkt. Wurden Teilnehmer ihres Experiments nach einem Slack-induzierenden Vertrag vergütetet, so schätzten diese am Ende der Budgetperiode Slack als unethisch ein. Bei den Teilnehmern, die einem wahrheitsinduzierendem System ausgesetzt wurden, war dies nicht der Fall. Aufgrund von Werten wie Empathie oder sozialen Normen ließe sich unter einem Slack-induzierenden System ein moralischer Rahmen schaffen, bei dem Manager zwischen Eigeninteresse und dem von ihnen erwarteten sozialen Verhalten abwägen. Dies legt nahe, dass die Ausprägung persönlicher Werte bestimme, wie einzelne Mitarbeiter auf veränderte Rahmenbedingungen reagieren (vgl. Hobson et al., 2011, S. 103 f.). Empfindet der Manager den Prozess der Anreizgestaltung als unfair, so führt dies zu einer stärkeren Generierung von Slack, jedoch nur wenn ein Slack-induzierender Vertrag zugrunde liegt. Bei einem wahrheitsinduzierenden Vertrag ist dies nicht der Fall. Empfundene Fairness spielt demnach nur bei Verträgen eine Rolle, die den Manager zur Verhandlung von Slack motivieren (vgl. Libby, 2015, S. 161). Diese Ergebnisse deuten darauf hin, dass eine faire Gestaltung des Anreizsystems ebenso wie moralische Werte unabhängig des gewählten Vertrages einen wichtigen Beitrag zur Gestaltung von Anreizsystemen leisten können (vgl. Hobson et al., 2011, S. 105).

Bei der Implementierung von Results Controls sollten Unternehmen v. a. bei Verwendung eines wahrheitsinduzierenden Anreizsystems auch die Risikoneigung ihrer Manager beachten. In einer Studie von Waller (1988) haben unter einem Slack-induzierenden Anreizsystem sowohl risikoaverse als auch risikoneutrale Agenten Slack in ihren Reports berichtet. Nachdem auf ein wahrheitsinduzierendes Schema gewechselt wurde, hat sich Slack bei den risikoneutralen Agenten signifikant verringert, bei den risikoaversen Agenten ließ sich jedoch keine Veränderung beobachten (vgl. Waller, 1988, S. 96).

Libby, Proell und Smith (2019) untersuchen, wie sich das Setzen schwer zu erreichender Ziele auf die Ehrlichkeit in Berichten von Managern auswirkt. Sie entlohnen die Manager dazu mit einem Fixgehalt und variieren den Grad der Entscheidungsfreiheit des Managers. Es zeigt sich, dass Manager mit einer hohen Zielvorgabe bei einem hohen Grad an Entscheidungsfreiheit ehrlicher berichten, als wenn diese fehlt. Daraus lässt sich schließen, dass in Unternehmen, die ein hohes Zielniveau in ihr MCS integrieren wollen, Manager auch mit den entsprechenden Entscheidungsfreiheiten ausgestattet werden sollten, um diese Ziele effektiv verfolgen zu können. Gleichzeitig verweisen die Autoren auf Vorsicht bei der Nutzung schwieriger Ziele, da diese in Verbindung mit starken finanziellen Anreizen auch zu einer verringerten Ehrlichkeit führen könnten (vgl. Libby et al., 2019, S. 196).

Im Rahmen der Budgetierung spielt auch die Investitionssteuerung eine Rolle. Hier werden neben Auszahlungen auch Einzahlungen betrachtet (vgl. Ewert & Wagenhofer, 2014, S.

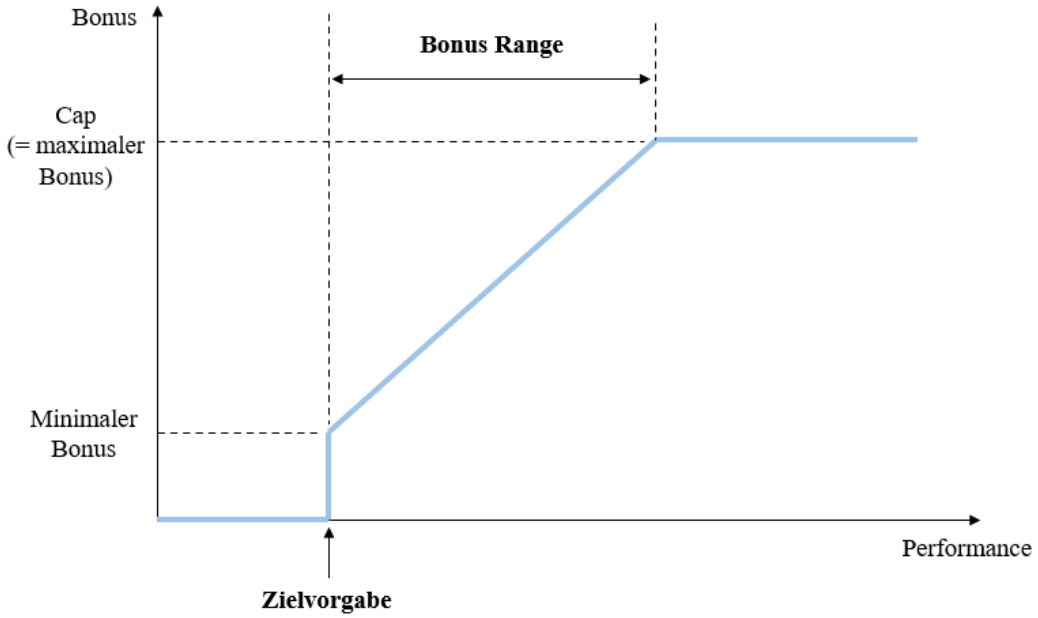

Abbildung 2: Budgetbasierter Hurdle- und Cap-Vertrag
Quelle: In Anlehnung an Fisher, Sprinkle & Walker, 2007, S. 70

455). Häufig wird dem Agenten dabei eine Hurdle-Rate vorgegeben, die einen Zielwert festlegt, ab welchem ein Projekt genehmigt wird. Durch das Festlegen einer Hurdle-Rate kann das Unternehmen anders als unter einem Trust-Kontrakt einen Teil der Überrendite abgreifen. In einer Kostenbetrachtung werden Berichte des Agenten dann akzeptiert, wenn die Kosten unter der Hurdle bleiben. Dem Agenten werden Mittel in Höhe der Hurdle-Rate zur Verfügung gestellt, die Differenz zwischen der Hurdle und den berichteten Kosten stellt dabei den „Gewinn" des Agenten dar. Liegen die berichteten Kosten über der Hurdle, wird das Projekt abgelehnt (vgl. Brink et al., 2018, S. 96). Dadurch könnten dem Unternehmen zwar profitable Projekte entgehen, gleichzeitig werden dadurch aber auch die Möglichkeiten des Agenten verringert, Slack auszubilden (vgl. Antle & Fellingham, 1995, S. 44). Aus Agency-Sicht berichtet ein Agent unter einer Hurdle-Rate seine tatsächlich erwarteten Kosten. Es wird jedoch angenommen, dass Hurdle-Rate-Kontrakte den Managern gegenüber Misstrauen ausdrücken und diese mit einem stärkeren unethischen Verhalten reagieren als dies unter einem Trust-Kontrakt der Fall wäre (vgl. Brink et al., 2018, S. 96).

Wird Managern sowohl unethisches Verhalten als auch eine Tendenz zur Ehrlichkeit unterstellt, so kann der Prinzipal die Vergabe von Budgets an die Berichte des Managers knüpfen, in das Budget aber gleichzeitig eine Hurdle einarbeiten. Die Hurdle kann dabei umso geringer ausfallen, je ehrlicher der Manager eingeschätzt wird. Die Nutzung von Hurdle-Rates kann es dem Unternehmen somit erlauben, einen Vertrag zu gestalten, der sich sowohl bei ehrlichen als auch bei unehrlichen Managern eignet (vgl. Mittendorf, 2006, S. 77 f.). Insgesamt betrachtet schlägt Mittendorf, 2006 zur optimalen Vertragsgestaltung vor, grundsätzlich den Berichten der Manager zu vertrauen, diese aber mit einer Hurdle-Komponente zu verbinden (vgl. Mittendorf, 2006, S. 90).

Ein klassischer Anreizvertrag ist wie in Abb. 2 dargestellt, ausgestaltet. Erreicht der Manager ein bestimmte Zielvorgabe, erhält er einen fixen Bonus. Gelingt es ihm nicht diese Hurdle (Zielvorgabe) zu schlagen, so erhält er keinen Bonus. Für jede zusätzliche Performance-Einheit über der Hurdle-Rate wird dem Manager ein bestimmter Anteil daran zugewiesen. Ab einem bestimmten Performance-Level wird der maximal erreichbare Bonus gedeckelt (Cap). Ab diesem Cap gibt es für den Agenten keinen finanziellen Anreiz mehr, zusätzliche Performance zu erzielen. Der variable Vergütungsanteil des Agenten bewegt sich dann innerhalb der Bonus-Range. Unter dem abgebildeten Budget-Kontrakt hat der Agent einen Anreiz ein möglichst geringes Budget zu melden, da dadurch unabhängig von der Performance sein Pay-Off maximiert wird. Entscheidend ist deshalb welches Performance-Level zwischen Manager und Prinzipal vereinbart wird (vgl. Fisher et al., 2007, S. 68).

Da bei der Nutzung von Hurdle-Rates in der Investitionssteuerung die Gefahr besteht, dass grundsätzlich profitable Projekte von Managern nicht umgesetzt werden, da sie die festgelegte Hurdle nicht überspringen, stellen Hurdle-Rates ein Control-Instrument dar, das mit hohen Kosten verbunden ist. Eine Nutzung erscheint deshalb nur sinnvoll, wenn ein hohes Maß an falscher Berichterstattung vermutet wird. Außerdem können Hurdles eine Signalwirkung auf den Manager ausüben, dass dessen Berichte von Unternehmensseite misstraut wird. Dieses Misstrauen kann wiederum dazu führen, dass ein Manager Informationen bewusst sabotiert. Dies kann sich dadurch äußern, dass Manager ein Projekt als nicht

lohnend kennzeichnen, obwohl das Projekt lohnend und damit umzusetzen wäre. Anders als es bei Slack der Fall ist, wo lediglich das Unternehmen an Profitabilität einbüßt, büßen in diesem Szenario beide Parteien an Profitabilität ein (vgl. Brunner & Ostermaier, 2019b, S. 72).

Nach Ansicht von Altenburger (2017) sollte kein Hurdle-Kontrakt verwendet werden. Demnach führe ein Trust-Kontrakt zu einer Maximierung der Wohlfahrt des Unternehmens, wobei das Unternehmen in seinen Überlegungen berücksichtigen sollte, welches Level an Ehrlichkeit wahrscheinlich erscheint. Zudem sollten die Unternehmen nicht vor mit der Implementierung eines Trust-Kontraktes verbundenen Kosten, wie z. B. Kosten für die Implementierung von sozialen Normen, zurückschrecken (vgl. Altenburger, 2017, S. 24).

Unternehmen können sich die Ehrlichkeitspräferenz von Individuen bei der Kontraktgestaltung zu Nutze machen. So kann das Vertrauen in die Ehrlichkeit des Managers zu einem höheren Unternehmensprofit führen als unter einem Vertrag mit einer Hurdle-Rate. Gleichzeitig kann das Vertrauen in die Ehrlichkeit aber auch zu einem geringeren Unternehmensprofit im Vergleich zur Hurdle-Rate führen. Als eine ex-ante-Variable, um das Ehrlichkeitsniveau der Manager einschätzen zu können, schlagen Blay, Douthit und Fulmer (2019) „Negative Affect Intensity“ vor. Darunter wird beschrieben, wie stark das Widerstreben eines Individuums ist, die sozialen Normen zu brechen und wie hoch der Disnutzen ist, den das Individuum dabei erleidet. Je höher die „Negative Affect Intensity“ ist, desto stärker ist die Präferenz zur Ehrlichkeit eines Managers und desto wirksamer erscheint ein Trust-Kontrakt (vgl. Blay et al., 2019, S. 63). Da es sich bei der „Negative Affect Intensity“ um eine ex-ante Variable handelt, können Unternehmen diese in ihre Überlegungen mit einbeziehen und in ihren Results Controls berücksichtigen.

Auch Jensen (2003) benennt die nicht linearen Bestandteile (Hurdle und Cap) eines klassischen Anreizsystems als hinderlich zur Implementierung einer wahrheitsgemäßen Berichterstattung. Er schlägt stattdessen vor, den Bonus linear von der Performance abhängig zu machen, wie in Abb. 3 dargestellt. Dadurch würde dem Manager der Anreiz zum Lügen entzogen werden. Dabei setzt er voraus, dass der Anreiz zum Lügen für den Manager am höchsten ist, wenn dieser seine Performance im Bereich einer der Schwellen werte vermutet (vgl. Jensen, 2003, S. 388 f.).

Zusammenfassend lässt sich festhalten, dass Manager durch die von ihnen abgegebenen Berichte einen großen Einfluss auf die vom Unternehmen festzusetzenden Ziele nehmen können. Je nach Vertragsausgestaltung ergibt sich dabei ein Anreiz zur Falschberichterstattung. Das Unternehmen hat in diesem Kontext verschiedene Ansatzpunkte, um sich die Ehrlichkeitspräferenzen ihrer Manager zunutze zu machen. Wichtig erscheint, dass sich Manager eine faire Vertragsgestaltung wünschen, Unternehmen gleichzeitig aber versuchen sollten, abzuschätzen, ob und welche Ehrlichkeitspräferenzen ihre Manager aufweisen.

Art der Belohnung

Der letzte Schritt in der Implementierung von Results Controls besteht in der Belohnung der Mitarbeiter. Grundsätzlich kann die Art der Belohnung dabei alles umfassen, was der Mitarbeiter wertschätzt (vgl. Merchant & Van der Stede, 2017, S. 40). Oftmals erfolgt die Anreizgestaltung in Form einer finanziellen Komponente (vgl. Chong & Wang, 2019, S. 279 f.). Neben finanziellen Belohnungen können Unternehmen ihre Mitarbeiter auch anhand von Sachbelohnungen wie bspw. Geschenkkarten oder Reisegutscheinen entlohnen (vgl. Presslee, Vance & Webb, 2013, S. 1827). Auch Belohnungen in Form von Lob sind denkbar (vgl. Lourenço, 2016, S. 283). Wird in der Forschung die Wirkung finanzieller Anreize untersucht, ist diese in Feldversuchen höher als in Laborexperimenten. Es ist wohl schwierig unter Laborbedingungen die komplexen Bedingungen der Realität zu modellieren (vgl. Garbers & Konradt, 2014, S. 119).

Die Wirkung (finanzieller) Anreize kann in zwei Bestandteile zerlegt werden. Zum einen führt die Art des Anreizes zu einem Motivationseffekt, der ein vom Unternehmen gewünschtes Verhalten für den Agenten attraktiver gestalten soll. Zum anderen bewirken Anreize einen psychologischen Effekt, welcher einen negativen Einfluss auf das gewünschte Verhalten nehmen kann. Tritt dieser psychologische Faktor auf, so wird vom Crowding-Out gesprochen, die Anreizgestaltung verdrängt dann durch den psychologischen Effekt die intrinsische Motivation zur Ehrlichkeit des Agenten (vgl. Gneezy, Meier & Rey-Biel, 2011, S. 192).

Mitarbeiter, die mit Sachgütern entlohnt werden, setzen sich selbst leichter zu erreichende Ziele, als dies bei einer Belohnung in finanzieller Form der Fall ist. Gleichzeitig zeigt sich aber auch, dass Mitarbeiter bei Sachbelohnungen stärker dazu motiviert sind, ihr selbst gesetztes Ziel auch tatsächlich zu erreichen. Die Endperformance betrachtend fällt das Ergebnis für das Unternehmen bei finanziellen Anreizen dennoch besser aus, aufgrund des höheren Levels an eingangs gesetzten Zielen (vgl. Presslee et al., 2013, S. 1827).

Oftmals erfolgt die Belohnung dabei in Form eines Bonus-Kontraktes, bei dem eine Mehrung des Wohlstandes des Managers erfolgt, wenn er sein Ziel erreicht. Daneben ist auch die Gestaltung als Penalty-Kontrakt möglich, bei dem Manager bestraft werden, sollten sie ein vorab festgelegtes Zielniveau nicht erfüllen können (vgl. Nichol, 2019, S. 329). Das Maß an unehrlicher Berichterstattung ist bei einem Penalty-Kontrakt höher als bei einem vergleichbaren Bonus-Kontrakt (vgl. Nichol, 2019, S. 342). Dies lässt sich mit Hilfe der Prospect-Theorie begründen, wonach Manager ein stärkeres Anspruchsempfinden auf den im Vertrag festgelegten Betrag entwickeln und im Zweifel auch durch Misreporting den Verlust des Betrages vermeiden wollen (vgl. Kahneman & Tversky, 1979, S. 330). Daraus lässt sich schließen, dass Penalty-Kontrakte zwar den Aufwand eines Agenten erhöhen können, gleichzeitig aber auch dessen Level an Unehrlichkeit, v. a. wenn der Agent droht an der Zielvorgabe zu scheitern (vgl. Nichol, 2019, S. 342). Entsprechend sollten Unternehmen diesen Punkt in der Gestaltung ihres Anreizsystems beachten. Gerade wenn das Performance-Ziel auf einem hohen Level ist und nicht absehbar ist, ob der Manager dieses Ziel erreichen wird, kann sich dies negativ

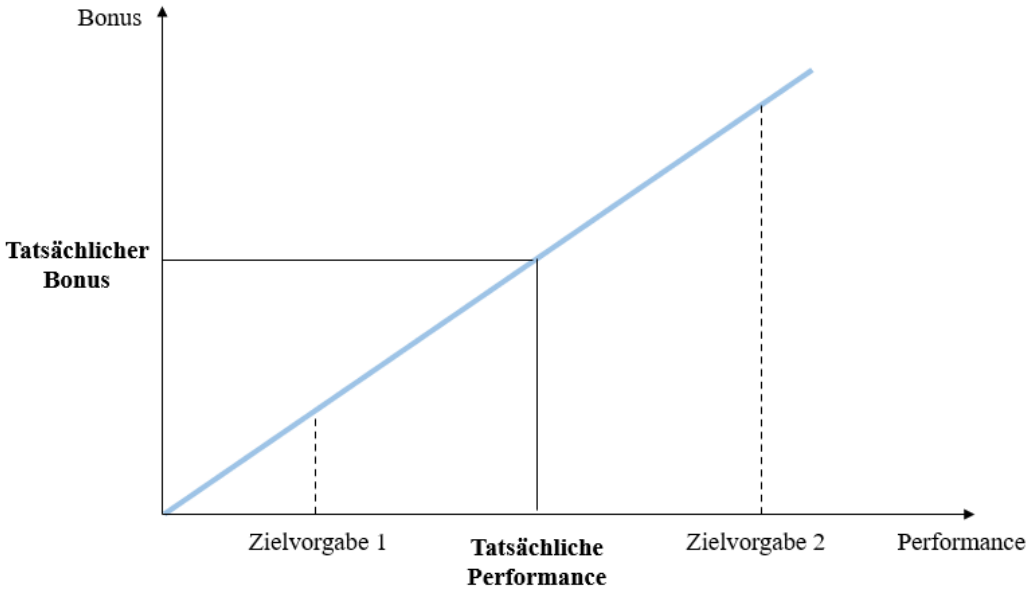

Abbildung 3: Linearer budgetbasierter Vertrag
Quelle: In Anlehnung an Jensen, 2003, S. 389

auf die wahrheitsgemäße Berichterstattung auswirken. Nichol (2019) schlägt dazu die folgende Unterscheidung vor: Wenn ein Unternehmen auf Bonus-Kontrakte setzt, so sollte es sich in seinem MCS darauf fokussieren, den Aufwand des Agenten zu erhöhen. Wird hingegen ein Penalty-Kontrakt genutzt, so sollte der Fokus der Steuerungsinstrumente auf einer Reduktion von Misreporting liegen (vgl. Nichol, 2019, S. 342).

Wie ehrlich Manager ihre privaten Informationen berichten, hängt auch davon ab, wie der durch den Manager erwirtschaftete Gewinn im Unternehmen verteilt wird. Demnach berichten Manager weniger ehrlich, wenn das von ihnen erwirtschaftete Vermögen in großen Teilen dem Unternehmen zugutekommt. Eine Möglichkeit sich die Ehrlichkeitspräferenz von Individuen zunutze zu machen, liegt für Unternehmen darin, statt einer variablen Bonuskomponente ein höheres Fixgehalt auszuzahlen, da dadurch der Anteil des Managers am Vermögen steigt (vgl. Evans et al., 2001, S. 553 f.).

Neben dem erwarteten Verhalten des Managers müssen Unternehmen auch ihre eigenen Rahmenbedingungen in das Anreizsystem miteinfließen lassen. So greifen bspw. verlustreiche Unternehmen, die das baldige Austreten eines Managers aus dem Unternehmen erwarten, oftmals auf nichtfinanzielle Anreize zurück. Dies ist darauf zurückzuführen, dass nichtfinanzielle Anreize zum einen eine motivierende Wirkung bezüglich langfristig verfolgter Ziele besitzen, außerdem sinkt so zum anderen der Anreiz des Managers durch bilanzpolitische Maßnahmen seine kurzfristige Position zu verbessern (vgl. Matějka, Merchant & Van der Stede, 2009, S. 903 f.).

Die Gestaltung eines Anreizsystems kann sich auch auf andere Agenten auswirken. So zeigen Cardinaels und Yin (2015), dass die Änderung eines Festgehalt-Kontraktes zu einem variablen Anreizsystem das Level an Misreporting bei anderen Agenten negativ beeinflussen kann. Die Veränderung des Anreizsystems verändert das Normverständnis der anderen Agenten und führt so zu einem höheren Level an Misreporting. Insgesamt werde so der Vorteil, den das Unternehmen aus der Veränderung im Anreizsystem ziehen wollte, wieder zunichte gemacht (vgl. Cardinaels & Yin, 2015, S. 1010).

Eine weitere Möglichkeit, um Mitarbeiter im Rahmen von Results Controls zu movieren, stellen Profit-Sharing-Pläne dar. Diese stellen eine Gewinnbeteiligung am Gewinn des Unternehmens dar. Das Ziel von Profit-Sharing-Plänen ist die Motivation von Mitarbeitern sowie die Verstärkung der Zielkongruenz zwischen den individuellen Zielen des Managers und den Unternehmenszielen. Gleichzeitig besteht bei Profit-Sharing die Gefahr, dass Manager dadurch weniger Vertrauen verspüren und somit ein Crowding-Out-Effekt eintritt, der in der Folge zu einer absinkenden Ehrlichkeit führt (vgl. Boster, Majerczyk & Tian, 2018, S. 697 f.). Boster et al. (2018) untersuchen deshalb die Auswirkungen von Profit-Sharing Plänen auf die Ehrlichkeit von Managern. Dabei unterscheiden sie zwischen individuellem Profit-Sharing, bei dem der Manager einen bestimmten Anteil am Gewinn seines Bereichs erhält, und gepooltem Profit-Sharing, wo der Manager einen Anteil des gesamten Unternehmensgewinns erhält. Kongruent zur Crowding-Out-Theorie kommen sie zu dem Ergebnis, dass unter einem individuellen Profit-Sharing-Plan weniger ehrlich berichtet wird als in einer Situation ohne Profit-Sharing-Plan. Das Profit-Sharing wird hier als Signaling-Instrument gesehen, welches das Misstrauen im Unternehmen fördert. Anders verhält es sich bei einem gepoolten Profit-Sharing. Die Autoren begründen dies mit einem erhöhten sozialen

Druck und dem Einfluss von Normen zur Ehrlichkeit, da Misreporting nun auch andere Agenten beeinflusst (vgl. Boster et al., 2018, S. 712). Weitere Aspekte von gruppenbasierten Belohnungen im Rahmen der Unternehmenssteuerung werden in Kapitel 4.4 näher behandelt.

3.3. Effektivität von Action Controls

Action Controls stellen eine sehr direkte Form von Steuerungsinstrumenten dar (vgl. Merchant & Van der Stede, 2017, S. 86). Sie beschreiben konkret, wie Handlungen von Managern ausgestaltet sein sollten, um deren Verhalten zu steuern (vgl. Malmi & Brown, 2008, S. 293). Im Rahmen von Action Controls werden Prozesse im Unternehmen sowie die Einhaltung regulatorischer Anforderungen überwacht, wodurch das Unternehmen vor Fehlern und betrügerischem Verhalten geschützt werden soll (vgl. Rae & Subramaniam, 2008, S. 105).

Zur effizienten Implementierung von Action Controls wird vorausgesetzt, dass das Unternehmen weiß, welche Handlungen wünschenswert sind und überprüfen kann, ob der Agent diese Handlungen auch tatsächlich im Sinne des Unternehmens durchgeführt hat (vgl. Feichter & Grabner, 2020, S. 162).

Merchant und Van der Stede (2017, S. 86) teilen Action Controls in vier verschiedene Arten ein: Verhaltensbeschränkungen, Pre-Actions Reviews, Action Accountability und Redundanz. Insbesondere zur Beseitigung von Motivational Problems können dabei alle vier Arten beitragen (vgl. Merchant & Van der Stede, 2017, S. 91 f.).

Verhaltensbeschränkungen teilen sich auf in physische (z. B. Schlösser, Zugangsodes) und administrative (z. B. Mitarbeiter darf nur über einen Geldbetrag bis 1.000€ eigenständig entscheiden) Beschränkungen (vgl. Merchant & Van der Stede, 2017, S. 86 f.). Diese Art der Steuerungsinstrumente wollen weniger durch eine Verhaltensänderung zum Ziel führen als vielmehr durch eine Limitierung der Handlungen des Managers (vgl. Feichter & Grabner, 2020, S. 162). Wird von Managern ausgegangen, die aufgrund von Informationsasymmetrien mit einer Vielzahl von Entscheidungsrechten ausgestattet sind (bspw. Länderverantwortliche in Konzernen) und mit ihren Berichten einen großen Verantwortungsbereich abdecken, erscheinen Verhaltensbeschränkung für deren Aufgabenerfüllung kontraproduktiv und als Steuerungsinstrument wenig geeignet. Folglich werden sie im Rahmen dieser Arbeit nicht näher betrachtet. Gleiches trifft auf die Redundanz zu, die durch die Zuweisung von mehr Mitarbeitern zu einer Aufgebe als benötigt gekennzeichnet ist (vgl. Merchant & Van der Stede, 2017, S. 91).

Eine andere Möglichkeit Action Controls zu unterscheiden, stellen präventive und detektive Controls dar. Präventive Controls ermöglichen es, Probleme bereits vor deren Entstehen zu erkennen, während detektive Controls zum Einsatz kommen, um Probleme zu erkennen, nachdem sie aufgetreten sind (vgl. Christ, Emett, Summers & Wood, 2012, S. 432). Beim Vergleich von präventiven und detektiven Controls zeigt sich, dass nicht die Anwendung der einzelnen Controls entscheidend zur Verringerung von unethischem Verhalten ist, sondern ob und zu welchem Zeitpunkt der Agent ein Feedback erhält. So können beide Instrumente dazu geeignet sein, das Verhalten des Agenten zu steuern, solange er ein unmittelbares Feedback erhält. Als Beispiel hierzu werden Maßnahmen einer Bank zur Bekämpfung von Korruption betrachtet. Ein denkbares präventives Control-Instrument wäre es hier, Überweisungen einzelner Mitarbeiter auf einen Höchstbetrag zu beschränken. Ein detektives Control-Instrument mit sofortigem Feedback könnte darin bestehen, dass ein Vorgesetzter eine Benachrichtigung erhält, wenn große Summen überwiesen werden. Bei einem detektiven Control-Instrument mit verzögertem Feedback könnte es sich um monatlich erstellte Übersichten mit großen Überweisungen handeln. Gleichzeitig ist hierbei anzumerken, dass durch präventive Controls ein Absinken der intrinsischen Motivation des Agenten beobachtet werden könnte, was wiederum zu einem verstärkten unethischen Verhalten führen könnte (vgl. Christ et al., 2012, S. 449).

3.4. Pre-Action Reviews in den Planungsprozessen

Aufgrund der Eigennutzorientierung des Agenten in der Agency-Theorie sollten Prinzipale grundsätzlich skeptisch den Agenten gegenüberstehen, was Maßnahmen der Überwachung notwendig macht (vgl. Helgesson, 2011, S. 473). Eine Möglichkeit bietet sich dabei in der Überprüfung von Plänen durch den Prinzipal, bevor diese genehmigt werden (vgl. Merchant & Van der Stede, 2017, S. 90).

Ob eine starke Beeinflussung durch das Unternehmen im Budgetprozess ein wirksames Steuerungsinstrument darstellt, erweist sich jedoch als zweifelhaft. So kann es zu einer Verringerung von Slack führen, wenn dem Agenten die Entscheidung über das Festlegen des finalen Budgets anstelle des Prinzipals überlassen wird. Dies kann durch ein stärkeres Maß an Ehrlichkeit der Agenten begründet werden. Außerdem kann eine Festsetzung des Budgets durch den Prinzipal zu so hohen Zielen führen, dass der Manager weniger Aufwand in die Erreichung der Ziele steckt. Unternehmen sollten deshalb hinterfragen, wie stark sie in die Abgabe des Budgets eingreifen wollen (vgl. Brown et al., 2017, S. 43).

Eine weitere Erklärung hierzu kann leisten, dass Individuen nur ungerne falsche Tatsachenberichte abgeben. Hat der Agent im Budgetprozess die letzte Entscheidung zu treffen und der Prinzipal fordert dazu eine bindende Erklärung von seinem Agenten an, so zeigt sich entsprechend, dass die Berichte weniger Slack enthalten. Dies deutet auf einen Einfluss von Ehrlichkeit im Berichtskontext hin, ohne dass weitere Anreize geschaffen worden wären (vgl. Rankin et al., 2008, S. 1095). Andererseits zeigt sich, dass das Abgeben einer Tatsachenerklärung keinen Einfluss auf die Slack-Bildung und das Reporting-Verhalten des Agenten nimmt, wenn der Prinzipal die letzte Entscheidung über das Budget zu treffen hat. Die Ehrlichkeitspräferenz des Agenten wirkt sich hier wohl nicht auf die Berichte aus. Rankin et al. (2008) argumentieren, dass eine solche Situation für den Agenten zu keinem ethischen Dilemma führt, sondern er diese Situation als Verhandlung einstuft, in der die Bildung von Slack akzeptiert

wird (vgl. Rankin et al., 2008, S. 1096). Douthit und Stevens (2015) kritisieren die Anwendbarkeit der Ergebnisse von Rankin et al. (2008) jedoch in realen Unternehmen. In der Studie hatten Manager Informationen über die relative Einkommensverteilung zwischen Prinzipal und Agent. Dies ist in der Realität in der Regel nicht der Fall. Aus den Ergebnissen der Studie sollte deshalb nicht auf die reale Situation in Unternehmen geschlossen werden (vgl. Douthit & Stevens, 2015, S. 487 f.).

Die Zuweisung und Wahrung von Partizipations- und Entscheidungsrechten kann dabei eine natürliche Möglichkeit darstellen, Slack zu verringern. V. a. langfristig sei laut Lau und Eggleton (2003) die Wahrung von Partizipationsrechten für Manager wichtig und wird von diesen gegenüber der kurzfristigen Erzielung von Slack bevorzugt. Basis für diese Überlegungen war das Eigennutzsterben der Manager. Langfristig gesehen könne demnach Misreporting nicht im Eigeninteresse des Managers stehen (vgl. Lau & Eggleton, 2003, S. 101).

Eine andere Art das Verhalten des Mitarbeiters im Rahmen von Pre-Action Reviews zu beeinflussen, kann die Kommunikation von Präferenzen des Unternehmens an seine Mitarbeiter darstellen. Dabei werden von Unternehmensseite lediglich die gewünschten Resultate kommuniziert, nicht aber welche Aktionen das Unternehmen plant, um diese Resultate zu erreichen. Diese Informationen werden nicht explizit genutzt, um sie in das Anreizsystem des Managers einzuarbeiten, was es zu einer indirekten Art von Steuerungsinstrumenten macht. Es zeigt sich, dass hier die Kommunikation spezifischer Präferenzen der Kommunikation allgemeiner Unternehmenspräferenzen bezüglich der Kosten vorzuziehen ist, um das Verhalten der Manager zu steuern. Doch bereits die Kommunikation von moderaten und unspezifischen Zielen kann die Ehrlichkeit erhöhen. Neben den genannten Vorteilen hat Newman (2014) in seiner Studie keine negativen Auswirkungen der Kommunikation von moderaten Kostenzielen feststellen können, was auf eine gute Eignung als Control-Instrument hindeutet (vgl. Newman, 2014, S.205).

Das Wording, in welchem Anfragen an den Agenten formuliert werden, spielt wohl keine Rolle zur Implementierung einer wahrheitsgemäßen Berichterstattung. So ließ sich kein Unterschied in der Slack-Bildung erkennen, nachdem das Budget auf verschieden formulierte Arten angefordert worden war (Anforderung eines ehrlichen, fairen oder vom Agenten präferierten Berichtes) (vgl. Brown et al., 2017, S. 43).

Die Abgabe von Berichten ist in der Praxis meist kein einmaliges, sondern ein periodisch wiederkehrendes Ereignis. Bei der Planung ihrer Budgets sollten Unternehmen deshalb auch den Zeithorizont berücksichtigen. So führt eine Situation, in der beide Parteien erwarten auch in der Zukunft wieder zusammen zu arbeiten, dazu, dass sich beide Parteien kooperativer in ihrem Verhalten zeigen. Zugleich zeigt sich auch, dass Manager, die zukünftige Verhandlungen mit demselben Prinzipal erwarten, eine bessere Performance erreichen als Manager, für die die Budgetverhandlung ein einmaliges Ereignis darstellt (vgl. Fisher, Frederickson & Peffer, 2006, S. 524).

Wird lediglich ein Prinzipal mit der Beaufsichtigung einer größeren Zahl von Agenten beauftragt, so kann dies aufgrund der Wirkung sozialer Normen auf Seiten des Prinzipals zu einem verringerten Misreporting führen. Ein Prinzipal neigt in dieser Situation aufgrund der ihm zur Verfügung stehenden Auswahlmöglichkeiten eher dazu Projekte abzulehnen, bei denen er eine hohe Gefahr der Slack-Bildung sieht. Die Agenten antizipieren dies und reagieren mit weniger Slack in ihren Berichten. Somit sinkt aus Unternehmenssicht im Durchschnitt der berichtete Slack pro Agent (vgl. hierzu und i. F. Hannan, Rankin & Towry, 2010, S. 526). Gleichzeitig weisen die Autoren darauf hin, dass sie im Rahmen ihrer Untersuchungen zum Thema sozialer Normen lediglich einen Teilaspekt untersuchen und in dieser Betrachtung andere Effekte auf die Organisation außen vorlassen. Demnach könnte der Gesamteinfluss auf das Unternehmen auch negativ sein. So bleibt zu erwähnen, dass eine höhere Anzahl von Agenten pro Prinzipal aufgrund der zugrunde liegenden Hidden Action Problematik auch zu einer geringeren Effizienz von einzelnen Steuerungsinstrumenten führen könnte.

Im Rahmen von Verhandlungen mit dem Prinzipal müssen sich auch die Agenten Informationen beschaffen. Dies vergrößert das Spektrum der Aktionen des Agenten. Noch bevor er einen Bericht erstellt und diesen an den Prinzipal weitergibt, steht der Manager vor der Entscheidung sich zusätzliche Informationen zu beschaffen oder dies zu unterlassen. Entscheidet sich der Agent dafür, vergrößert er damit die Informationsasymmetrie zwischen ihm und dem Prinzipal. Dieser angestiegene Informationsvorsprung führt jedoch nicht zu einem absolut höheren Level an Misreporting. Stattdessen wird durch die Charakterausprägungen des Agenten definiert, ob er diese zusätzlichen Informationen für weiteres opportunistisches Verhalten nutzt. Eine Besonderheit gibt es hier bei Agenten, die auf der einen Seite einen persönlichen Nutzen aus den Berichten an den Prinzipal ziehen wollen und sich zum anderen unwohl damit fühlen Informationen nicht wahrheitsgemäß zu berichten. Diese Gruppe von Agenten unterlässt eine zusätzliche Informationsbeschaffung und berichtet zeitglich unehrlicher an den Prinzipal, um dieses Dilemma zu umgehen (vgl. Church, Lynn Hannan & Kuang, 2014, S. 415).

Wenn sich das Unternehmen zur Nutzung von Pre-Action Reviews entscheidet, sollten die Prozesse klar definiert und verbindlich gestaltet werden. Ansonsten kann sich ein Problem dadurch ergeben, dass die Manager zwar formal einem Steuerungssystem unterliegen, die Controls dessen aber vermeiden können und somit die Möglichkeit haben sich auch weiterhin unethisch zu verhalten. Gino, Krupka und Weber (2013) haben diese Situation untersucht und den Studienteilnehmern die Entscheidung überlassen, ob ihre Berichte einer freiwilligen Kontrolle unterzogen werden. Dabei zeigt sich in der Gruppe derer, die sich der Kontrolle unterziehen sowohl ein höherer Grad an Misreporting als auch eine geringere Performance im Vergleich zu einer Gruppe ohne freiwillige Kontrolle. Dies lässt darauf schließen, dass sich bei den

Individuen ein höherer Grad an Unehrlichkeit einstellt, wenn sie sich bewusst dafür entscheiden können (vgl. Gino et al. (2013), S. 2200 f.).

3.5. Zuweisung von Verantwortlichkeiten

Ein weiterer Ansatzpunkt, um das Verhalten von Mitarbeitern im Rahmen von Action Controls zu steuern ist die Zuweisung von Verantwortlichkeiten für deren Handlungen (vgl. Merchant & Van der Stede, 2017, S. 90). Eine Möglichkeit dafür stellen Audits dar. Laut Arnold und Schreiber (2013) verringern Audits die Generierung von Slack und erhöhen folglich den Pay-Off des Prinzipals unter der Bedingung, dass dieser über mehrere Perioden hinweg immer wieder tauscht und der Agent nicht abschätzen kann, welchen Vorgesetzten er in der Folgeperiode haben wird. Wenn der Prinzipal in jeder Periode gleichbleibt, lässt sich keine Slack-Reduzierung feststellen. Demnach würde hier die Effektivität von Audits darunter leiden, dass der Agent sein Verhalten an das Verhalten des Prinzipals anpasst und ein zu striktes Verhalten von Seiten des Unternehmens zu einer vermehrten Slack-Bildung in Folgeperioden führe (vgl. Arnold & Schreiber, 2013, S. 208). Das Durchführen von Audits kann ein Unternehmen dabei unterstützen Misreporting zu erkennen und in der Folge zu einem geringeren Level an Misreporting im Unternehmen beizutragen. Audits sind dabei auch eng mit dem effektiven Einsatz von sozialen Normen verknüpft. Demnach reagieren Manager stärker auf die im Unternehmen vorherrschenden sozialen Normen, wenn das Unternehmen Audits durchführt. Dies ist darauf zurückzuführen, dass Audits das Bewusstsein von Managern stärken, welche sozialen Normen im Unternehmen wichtig sind. In der Folge erhöhen sich die Kosten der Lüge, wodurch sich auch die Entscheidung eines Managers nochmals ändern kann, der zuvor nicht wahrheitsgemäß berichten wollte (vgl. Cardinaels & Jia, 2016, S. 200). Um Reputationseffekte zu vermeiden, könnte es eine Möglichkeit darstellen Audits von Drittparteien durchführen zu lassen, auch eine unregelmäßige stichprobenartige Durchführung von Audits erscheint möglich. Bezüglich der Prinzipal-Agenten-Beziehung scheinen jedoch die Vorteile aus einer persönlichen Beziehung zwischen beiden Parteien nicht auszureichen, um im Rahmen von Audits eine Vertrauensbasis zu schaffen, in der die Bildung von Slack verringert wird (vgl. Arnold & Schreiber, 2013, S. 208 f.).

Brown, Fisher, Sooy und Sprinkle (2014) untersuchen, wie sich Rankings auf die Ehrlichkeit von Managern auswirken. Dazu sollen in ihrem Experiment Studienteilnehmer ihre Projektkosten unter dem Anreiz zum Misreporting wahrheitsgemäß an den Prinzipal berichten und werden anschließend nach ihrer Vergütung oder nach dem Anteil am Gewinn des Unternehmens beurteilt und ein Ranking dazu erstellt. Dabei hat sich gezeigt, dass ein Ranking auf Grundlage des Unternehmensgewinns die Ehrlichkeit der Studienteilnehmer erhöht hat, während bei dem Ranking auf Basis der Entlohnung genau das Gegenteil der Fall war. Die Veröffentlichung von Rankings könnte demnach von Unternehmen als ein kostengünstiges Control-Instrument eingesetzt werden. Dabei sollten die Unternehmen jedoch beachten, dass nicht jedes Ranking zu mehr Ehrlichkeit im Unternehmen führt, und genau entscheiden auf welcher Grundlage sie ein Ranking für ihre Mitarbeiter erstellen wollen (vgl. Brown et al., 2014, S. 244 f.).

Eine Möglichkeit Manager zu einer wahrheitsgemäßen Berichterstattung zu motivieren, kann es auch darstellen, diese zu sensibilisieren, wie ihre Entscheidungen die Gesamtsituation im Unternehmen beeinflussen. Zeigt der Bericht eines Managers auf, ob das Unternehmen durch die Tätigkeiten insgesamt einen Gewinn oder Verlust erzielt, so enthält der Bericht weniger Slack im Vergleich zu einem Bericht, in dem diese Komponente fehlt, aber auch im Vergleich zu einem Bericht, in dem die Entscheidungen des Managers keinen Einfluss auf das Unternehmensergebnis nehmen, d. h. das Unternehmen erzielt unabhängig davon einen Gewinn oder Verlust. Dies induziert, dass es sich positiv auf die Ehrlichkeit von Managern auswirkt, wenn diesen bewusst gemacht wird, dass ihr Verhalten einen Einfluss auf die Unternehmenssituation nimmt. Zumindest unter Experimentalbedingungen hätten einige Manager ihre Kosten sogar bewusst unterschätzt, um damit zur Profitabilität des Unternehmens beizutragen (vgl. Cardinaels, 2016, S. 60). Der verstärkte Einsatz von Steuerungsinstrumenten ist in diesem Kontext kritisch zu sehen. Erzielt ein Unternehmen Verluste, neigen Unternehmen tendenziell dazu verstärkt Control-Instrumente einzusetzen, um das Verhalten ihrer Mitarbeiter zu steuern. Die Erkenntnisse von Cardinaels (2016) betrachtend, könnte die Situation allein Manager bereits zu einer wahrheitsgemäßen Berichterstattung motivieren und zusätzliche Steuerungsinstrumente kontraproduktiv wirken (vgl. Cardinaels, 2016, S. 61).

Im Rahmen der Aufgabendelegation übertragen Unternehmen immer auch einen gewissen Einfluss an die Manager. Dies kommt einem psychologischen Vertrag gleich. Menschen neigen jedoch dazu den Einfluss, den sie tatsächlich haben, zu überschätzen. Wird die dadurch entwickelte Erwartungshaltung des Managers zu einem späteren Zeitpunkt enttäuscht, stellt dies für diesen einen psychologischen Vertragsbruch dar, worauf der Manager mit einer stärkeren Tendenz zu unethischem Verhalten reagiert. Eine große Gefahr besteht dann für das Unternehmen darin, dass der Manager diesen empfunden Vertragsbruch nicht an das Unternehmen kommuniziert, sondern seinen Level an Misreporting erhöht, ohne dass dies vom Unternehmen bemerkt wird. Unternehmen sollten folglich klar festlegen, welchen Einfluss ihre Mitarbeiter auf Entscheidungen und Prozesse nehmen dürfen und dies entsprechend klar an die verantwortlichen Mitarbeiter kommunizieren (vgl. Gallani, Krishnan, Marinich & Shields, 2019, S. 2940 f.).

Auch die Anwendung von Codes of Conduct kann ein Steuerungsinstrument im Sinne von Action Controls darstellen (vgl. Merchant & Van der Stede, 2017, S. 90), aufgrund der engen Verzahnung dieser mit informellen Controls werden Codes of Conduct in Kapitel 4.3 betrachtet.

4. Informelle Steuerungsinstrumente zur Sicherstellung einer wahrheitsgemäßen Berichterstattung

4.1. Effektivität von informellen Controls

Informelle Controls werden von Merchant und Van der Stede (2017, S. 95-97) in Personnel Controls und Cultural Controls unterschieden. Anders als formelle Controls werden informelle Controls benutzt, um das Verhalten der Mitarbeiter ex-ante zu beeinflussen (vgl. Voss & Brettel, 2014, S. 571 f.). Dabei beruht die Effektivität von informellen Controls auf Verpflichtungen und sozialem Druck (vgl. Lange, 2008, S. 712). Wirkt ein höherer sozialer Druck auf einen Manager, so führt dies zu einer geringeren Meldung von Slack (vgl. Young, 1985, S. 840). So zeigt sich, dass bereits das Wissen, was die anderen Mitglieder der Peer-Group erwarten, ausreicht, um das Reporting-Verhalten einzelner Personen zu verändern (vgl. Altenburger, 2017, S. 24). Die Beachtung sozialer Normen kann demnach eine wichtige Rolle in der Implementierung eines wirksamen MCS zur Vermeidung von Misreporting spielen.

Personnel Controls bauen auf einem Self-Monitoring-Ansatz auf, der auf der Eigenmotivation der Mitarbeiter fußt. Dazu soll Mitarbeitern verständlich gemacht werden, was das Unternehmen von ihnen will und die Mitarbeiter sollen mit den notwendigen Eigenschaften ausgestattet werden, um die Ziele des Unternehmens erreichen zu können (vgl. Feichter & Grabner, 2020, S. 163). Eine Implementierung von Personnel Controls kann durch die Selektion von Mitarbeitern, der Weiterbildung von Mitarbeitern oder durch Arbeitsplatzgestaltungen erfolgen (vgl. Merchant & Van der Stede, 2017, S. 95).

Cultural Controls knüpfen an die Unternehmenskultur an, die sich durch die Gesamtheit der im Unternehmen vorherrschenden Werte und Normen ergibt (vgl. Malmi & Brown, 2008, S. 294). Dabei sollen durch ein Mutual-Monitoring die Mitarbeiter gegenseitig durch sozialen Druck das Verhalten anderer Mitarbeiter beeinflussen (vgl. Feichter & Grabner, 2020, S. 162). In den folgenden Kapiteln werden die Cultural Controls Codes of Conduct, gruppenbasierte Belohnungen, Gestaltung des Arbeitsumfeldes sowie Tone at the Top betrachtet (vgl. Merchant & Van der Stede, 2017, S. 98).

Um informelle Controls effektiv umzusetzen, sollte ein Bewusstsein für gewünschte Ziele und erstrebenswertes Verhalten im Unternehmen geschaffen werden. So nehmen soziale Gruppennormen v. a. dann einen Einfluss, wenn ein Manager Prüfungen unterliegt, ob er in seinem Verhalten diesen Normen auch tatsächlich folgt (vgl. Cardinaels & Jia, 2016, S. 200 f.). Entscheidungen bezüglich Steuerungsinstrumenten haben oft eine unmittelbar sichtbare Konsequenz, wie es bspw. der Fall ist, wenn ein Unternehmen sein Vergütungssystem ändert. Gleichzeitig kann das Unternehmen mit informellen Controls auf einer unterbewussten Ebene agieren, auf der Veränderungen zwar nicht sofort sichtbar sind, gleichzeitig aber dennoch einen wichtigen Einfluss auf das Verhalten der Mitarbeiter ausüben (vgl. Cardinaels & Yin, 2015, S. 986). Soziale Normen bleiben im Zeitablauf stabil. Deshalb kann es sich für Unternehmen lohnen, in die Schaffung bzw. die Modifikation von sozialen Normen zu investieren, um dadurch zu einer wahrheitsgemäßen Berichterstattung beizutragen. Ein besonderes Augenmerk sollten Unternehmen auf Abweichler von den gültigen sozialen Normen legen, da diese sonst langfristig der Ehrlichkeit der anderen Mitarbeiter schaden könnten (vgl. Altenburger, 2017, S. 24).

Informelle Controls sind besonders gut geeignet, um Lack of Direction und Personal Limitations zu adressieren. Auch wenn Mitarbeiterselektion und gruppenbasierte Entlohnungen Steuerungsinstrumente darstellen, die lediglich direkt auf die Behebung von Motivational Problems einwirken (vgl. Merchant & Van der Stede, 2017, S. 101), erscheint es einleuchtend, dass durch wirkungsvoll gestaltete informelle Controls das Auftreten von Motivational Problems verhindert werden kann.

4.2. Selektion und Weiterbildung der Mitarbeiter

Abernethy und Brownell (1997) untersuchen basierend auf einem Modell von Perrow (1970), welche Arten von Steuerungsinstrumenten sich am besten für Aufgaben eignen, die keiner Routine unterliegen. Routine wird dabei definiert durch die Analysierbarkeit der Aufgaben sowie der Anzahl an Ausnahmen. Sie merken an, dass sich Personnel Controls zur Steuerung von Mitarbeitern umso besser eignen, je weniger Routine diese in ihren Aufgaben unterliegen (vgl. Abernethy & Brownell, 1997, S. 245). Für verantwortungsvolle Tätigkeiten mit einer hohen Entscheidungsfreiheit scheinen Personnel Controls entsprechend wirkungsvolle Steuerungsinstrumente darzustellen. Eine effektive Auswahl und Weiterbildung von Mitarbeitern sorgt für einen geeigneten Pool an Mitarbeitern, die in der Lage sind, zur Erreichung der Unternehmensziele beizutragen (vgl. Abernethy & Brownell, 1997, S. 239).

Die Nutzung von Personnel Controls kann die Anfälligkeit für opportunistisches Verhalten unter den Mitarbeitern abschwächen und dadurch die Anzahl notwendiger weiterer Steuerungsinstrumente verringern (vgl. Voss & Brettel, 2014, S. 581). Personnel Controls können zudem als komplementär zu Results Controls angesehen werden. So kann die Einstellung passender Mitarbeiter dafür sorgen, dass diese Mitarbeiter im Anreizsystem des Unternehmens eine gute Performance liefern (vgl. Abernethy, Dekker & Schulz, 2015, S. 659).

Die richtige Selektion von Mitarbeitern kann einen wichtigen Beitrag zur Zielkongruenz zwischen Prinzipal und Agenten leisten (vgl. Liao, 2005, S. 297). Dies betont die Wichtigkeit im Rahmen von Einstellungsverfahren die richtigen Mitarbeiter auszuwählen. Das Durchführen von Checks vor der Einstellung kann dem Unternehmen eine erste Indikation geben, ob es sich um einen ehrlichen oder einen unehrlichen Manager handelt. Gleichzeitig sollte das Ergebnis solcher Checks lediglich als Richtwert angesehen werden, da auch grundsätzlich ehrliche Manager unter den falschen Rahmenbedingungen zu unethischen Handlungen tendieren können (vgl. Chong & Wang, 2019, S. 299). Eine Möglichkeit solcher Tests stellen z. B. Integritätschecks dar, bei denen Bewerber mit einem hohen Level an Unehrlichkeit bereits vor

Einstellung erkannt werden sollen (vgl. Berry, Sackett & Wiemann, 2007, S. 272). Neben Tests können Unternehmen auch im Rahmen von Interviews auf ein Screening setzen und versuchen herauszufinden, welche Bewerber sich im späteren Berufsleben ethisch verhalten würden (vgl. Cleek & Leonard, 1998, S. 628).

Reziprozität stellt eine soziale Norm dar. In dieser geht es darum, dass Menschen anderen Menschen helfen, die diesen zuvor geholfen haben. Außerdem sollten Menschen den Menschen, von welchen sie Hilfe erfahren haben, keinen Schaden zufügen (vgl. Davidson, 2019, S. 87). Dadurch kann Reziprozität auch eine Anwendung im MCS eines Unternehmens finden, wenn die Einstellung eines Mitarbeiters bereits aktiv durch den Prinzipal begleitet wird. Wird ein Manager durch den Prinzipal persönlich eingestellt (z. B. in Form von Vorstandsmitgliedern), so zeigt sich in der Folge ein höheres Ehrlichkeitsniveau beim Manager. Dies lässt sich als eine Antwort des Managers auf das vom Prinzipal entgegen gebrachte Vertrauen interpretieren (vgl. Davidson, 2019, S. 98 f.).

Neben der Einstellung der richtigen Mitarbeiter stellen auch Weiterbildungsmaßnahmen eine Möglichkeit dar, Manager zu einer wahrheitsgemäßen Berichterstattung zu motivieren. Im Rahmen von Trainings können Mitarbeitern Werte vermittelt werden, die aus Unternehmenssicht wünschenswert sind. Dies kann dazu beitragen, dass sich Mitarbeiter v. a. in ethischen Dilemmas für die „richtigen" Handlungen entscheiden (vgl. Warren, Gaspar & Laufer, 2014, S. 89). Trainings sollen die Mitarbeiter dazu mit dem nötigen Wissen ausstatten und können so einen Beitrag zu einem ethischeren Arbeitsklima leisten (vgl. Tenbrunsel, Smith-Crowe & Umphress, 2003, S. 290). Mayhew und Murphy (2009) untersuchen an Studierenden, wie sich Weiterbildungen im Bereich der Ethik auf deren Berichtsverhalten auswirken. Dabei stellen sie fest, dass sich das Misreporting der Studienteilnehmer nicht verändert, wenn deren Berichte nicht veröffentlicht werden. Werden die Berichte jedoch veröffentlicht, so zeigt sich, dass der Grad an Misreporting in Folge der Weiterbildung abnimmt. Die Vermittlung ethischer Werte an Mitarbeiter könnte folglich das ethische Klima im Unternehmen verbessern. Die Studie legt aber nahe, dass eine bloße Wissensvermittlung über gewünschtes Verhalten nicht ausreicht, sondern dass das Unternehmen auch einen gewissen sozialen Druck als Komponente miteinbeziehen sollte (vgl. Mayhew & Murphy, 2009, S. 409).

Auch Konkurrenz ist ein Faktor, der Einfluss auf das Misreporting nehmen kann. Dabei kann Konkurrenz in dem Sinne verstanden werden, dass mehrere Mitarbeiter auf das gleiche Ziel hinarbeiten. Erreicht ein Mitarbeiter das Ziel, so scheitern die anderen Manager an der Erreichung dieses Ziels (vgl. Schreck, 2015, S. 178). In Unternehmen konkurrieren oftmals mehrere Manager untereinander um die Vergabe von Budgetmitteln. Bei einer Untersuchung mit unterschiedlicher Ausprägung des Konkurrenzdrucks (niedrig/mittel/hoch) zeigt sich bei Brüggen und Luft (2011), dass der Grad an Misreporting bei einem mittleren Konkurrenzdruck am höchsten ist und sowohl bei niedrigem als auch bei hohem Konkurrenzdruck abnimmt. Dies liegt an Managern, die nur eine geringe Rendite erwarten, mit ihren Berichten aber auf die Zuteilung von Mitteln hoffen. Während die Manager bei niedrigem Wettbewerb auf eine Zuteilung der Mittel auch bei ehrlicher Berichterstattung hoffen können und bei starker Konkurrenz die Wahrscheinlichkeit die Mittel zu erhalten selbst bei Misreporting gering ist, hat der Manager bei der mittleren Konkurrenzsituation den Anreiz mit seiner Lüge mit einer hohen Wahrscheinlichkeit die gewünschten Mittel zugewiesen zu bekommen (vgl. Brüggen & Luft, 2011, S. 409 f.). Neben Vorteilen, die Konkurrenz mit sich bringt (z. B. erhöhte Motivation, erleichterte Kapitalzuteilung), kann eine unpassende Ausgestaltung der Organisationsstruktur demnach zur falschen Weitergabe privater Informationen führen (vgl. Schreck, 2015, S. 85 f.). Unternehmen sollten dies bedenken, wenn sie sich dazu entscheiden wie viele Mitarbeiter um bestimmte Mittel konkurrieren.

Einen weiteren Ansatzpunkt für die Unternehmenssteuerung stellen die Mitarbeiter der Manager dar. Oftmals sind Manager zur Erstellung der eigenen Berichte auf den Input ihrer Mitarbeiter angewiesen. Entsprechend kann Misreporting bereits verzerrte Informationen von Seiten der Mitarbeiter erfordern. Damit könnten Mitarbeiter der Manager als informelles Control-Instrument dienen, indem diese das opportunistische Verhalten ihrer Vorgesetzten nicht unterstützen (vgl. Jollineau, Vance & Webb, 2012, S. 1 f.). Dabei steht weniger das aktive Aufdecken des unethischen Verhaltens des Managers im Vordergrund als vielmehr präventive und subtile Methoden, mit denen sich Mitarbeiter dem Verhalten ihres Vorgesetzten widersetzen können. In einer Studie von Jollineau et al. (2012) führte eine als gut eingeschätzte Beziehung zwischen Mitarbeiter und Manager dazu, dass der Mitarbeiter eher dazu bereit war, das Verhalten des Managers zu unterstützen und selbst verzerrte Informationen abzugeben. Dies war selbst dann der Fall, wenn sich dadurch die erwartete Vergütung des Mitarbeiters verringerte. Die Autoren der Studie verweisen explizit darauf, dass Unternehmen nicht versuchen sollten, eine schlechte Beziehung zwischen Angestellten zu schaffen, da dies eine Vielzahl von negativen Effekten nach sich ziehen könnte. Dennoch sollten sich Unternehmen bei der Gestaltung ihres MCS dieser Problematik bewusst sein. (vgl. Jollineau et al., 2012, S. 18 f.). Im Unternehmen könnte die Schaffung eines organisationalen Rahmens helfen, der Whistleblowing vereinfacht, um die Mitarbeiter zur Weitergabe von beobachtetem unethischem Verhalten zu motivieren und diese zugleich zu schützen (vgl. Gundlach, Douglas & Martinko, 2003, S. 119). Möglichkeiten bieten sich hier z. B. in der Einrichtung einer unternehmensexternen Hotline, um Verstöße anonym melden zu können (vgl. Kaptein, 20111, S. 527).

4.3. Codes of Conduct als Leitbild im Unternehmen

Codes of Conduct entsprechen einem Verhaltenskodex im Unternehmen, der eine Sammlung von Richtlinien enthält und sich an alle Mitarbeiter im Unternehmen wendet (vgl. Kreipl, 2020, S. 187). Sie erfüllen den Zweck, das Verhalten von Mitarbeitern zu steuern und sollen damit das Unterneh-

men vor illegalem und schädlichem Verhalten schützen (vgl. Adams, Tashchian & Shore, 2001, S. 199).

Studien zeigen, dass unethisches Verhalten in Unternehmen mit Codes of Conduct weniger stark verbreitet ist (vgl. Somers, 2001, S. 192). Unternehmen, deren Codes of Conduct eine hohe Qualität aufweisen, werden auch in Rankings bezüglich ihrer ethischen Performance hoch gelistet. Eine Verbindung der Codequalität zur Effektivität dieser liegt nahe (vgl. Kaptein, 2004, S. 544 f.).

Um Codes of Conduct möglichst effizient im Unternehmen einsetzen zu können, sollte im Unternehmen ein Bewusstsein für die vorherrschenden Codes sowie den Zielen, die damit erreicht werden sollen, geschaffen werden. Dadurch, dass Manager den Inhalt der Codes of Conduct kennen, lässt sich möglicherweise auch ein positiver Effekt darauf erzielen, als wie ethisch die Manager ihr Unternehmen im Gesamten einschätzen (vgl. Wotruba, Chonko & Loe, 2001, S. 66). Das Vorhandensein von Codes of Conduct im Unternehmen beeinflusste die Wahrnehmung von unethischem Verhalten im Unternehmen in einer Studie von Somers (2001) nicht. Hier zeigte sich jedoch Aufholbedarf auf Seiten der Kommunikation von Codes of Conduct im Unternehmen. In einer Befragung von 613 Management Accountants in verschiedenen US-Unternehmen konnten lediglich acht Prozent der befragten Angestellten mit Sicherheit sagen, dass ihr Unternehmen Codes of Conduct implementiert hatte. Statistiken aus dieser Zeit zeigen, dass jedoch ca. 75 Prozent der US-Unternehmen Codes of Conduct angewandt hätten (vgl. Somers, 2001, S. 192).

Chonko, Wotruba und Loe (2003) argumentieren, dass ein Code of Conduct im Unternehmen umso wirksamer ist, je genauer die Manager den Inhalt und die Ziele des Codes kennen. Die Autoren verweisen auf frühere Studien, die zu dem Ergebnis kamen, dass die Existenz eines Codes nicht ausreiche, um für ein ethisches Arbeitsklima zu sorgen. Demnach scheitern diese Studien daran, dass nicht die Existenz eines Codes an sich ausreicht, sondern dass die Mitarbeiter mit den Inhalten der Codes vertraut sein sollten, um dadurch in ihrem Verhalten beeinflusst zu werden (vgl. Chonko et al., 2003, S. 248).

Jedoch kann auch bereits das Vorhandensein von Codes of Conduct das Verhalten von Mitarbeitern positiv beeinflussen. So folgen Adams et al., 2001 der Position, dass das Vorhandensein eines Codes viel wichtiger sei als der genaue Inhalt dessen. Demnach könnten die meisten Mitarbeiter den Inhalt des in ihrem Unternehmen gültigen Codes nicht wiedergeben, gleichzeitig haben diese aber eine wesentlich andere Wahrnehmung von Ethik als Mitarbeiter in Unternehmen, die keinen Code of Conduct nutzen. Entsprechend sei nicht der Inhalt, sondern die Existenz sowie die Kommunikation dieser im Unternehmen wichtig (vgl. Adams et al., 2001, S. 207).

Haesebrouck (2021) untersucht, wie sich die Berichterstattung in einem Unternehmen, in dem Ehrlichkeit einen hohen Stellenwert besitzt, in Verbindung mit einem hohen Aufwand der Informationsbeschaffung verhält. Demnach enthalten Berichte von Managern mehr Slack, wenn mit der Berichterstellung ein hoher Aufwand der Informationsbeschaffung verbunden ist und zugleich Ehrlichkeit im Unternehmen keinen hohen Stellenwert einnimmt. Manager sehen in dieser Situation Slack als eine Form der Kompensation für den von ihnen betriebenen Aufwand an. Anders verhält es sich, wenn der Stellenwert von Ehrlichkeit im Unternehmen hoch ist. Dann zeigt sich, dass die Manager trotz des gesteigerten Aufwands ehrlicher berichten. Gemäß der Ownership-Theorie liegt der Grund dafür in einem gesteigerten Verantwortungsbewusstsein in dieser Situation. Wenn sich Unternehmen in einem Berichtsumfeld wiederfinden, in dem der Aufwand zur Informationsgewinnung groß ist, kann folglich die Förderung von Ehrlichkeit in Form von Steuerungsinstrumenten wie eines Code of Conduct einen Beitrag dazu leisten, dass unethisches Verhalten vermieden wird (vgl. Haesebrouck, 2021, S. 3107 f.).

Codes of Conduct können auch zu einer verbesserten Teamidentität beitragen. Von einem horizontalen Anreizsystem wird gesprochen, wenn Manager direkten Einfluss auf Aktionen anderer Manager ausüben können (vgl. Towry, 2003, S. 1070). Eine starke Teamidentität verstärkt die Kooperation zwischen Managern und erhöht damit ihren Aufwand. Manager erfahren dadurch folglich eine gegenseitige Motivation, was sich in einer wahrheitsgemäßen Berichterstattung äußern kann (vgl. Towry, 2003, S. 1088).

Von ihrer Persönlichkeit eher individualistisch veranlagte Manager neigen verstärkt zur Falschberichterstattung. Besonders stark negativ wirken sich dabei auf diese Gruppe von Menschen einschränkende Arten von Controls aus, wie z. B. negativ formulierte Codes of Conduct. Dies lässt sich darauf zurückführen, dass diese Menschen auf einschränkende Maßnahmen psychologisch mit dem Misreporting ihrer privaten Informationen reagieren und weiterhin versuchen ihre Interessen durchzusetzen. Sind die Codes positiv formuliert und deuten damit eher auf kommunizierte Leitbilder und Werte des Unternehmens hin, ist dieser Effekt nicht so stark ausgeprägt (vgl. Sánchez-Expósito & Naranjo-Gil, 2017, S. 588 f.).

Während die meisten Studien darauf hinweisen, dass eine negative Formulierung in Codes of Conduct möglichst vermieden werden sollte, da dies das Verhalten möglicherweise weniger effektiv beeinflussen könnte, findet Schwartz (2004) in vom ihm durchgeführten Interviews heraus, dass sich Mitarbeiter negativ formulierte Codes wünschen würden, da sie diese als effektiver und leichter umzusetzen einschätzen. Positiv formulierte Codes könnten demnach Mitarbeitern weniger klare Handlungsanweisungen mit an die Hand geben und diesen damit einen größeren Freiraum zum Misreporting einräumen (vgl. Schwartz, 2004, S. 329). Für die effektive Implementierung sollten Beispiele mit in den Code eingearbeitet werden und negative Formulierungen für explizit unerwünschtes Verhalten gewählt werden (z. B. „Du darfst keine Geschenke von Kunden annehmen.“). Außerdem sollten die Mitarbeiter in die Erstellung der Codes miteingebunden werden, um sicherzustellen, dass die enthaltenen Richtlinien in der Praxis auch wirklich Relevanz aufweisen und umzusetzen sind (vgl. Schwartz, 2004, S. 339). Ansonsten können

Codes of Conduct auch negative Auswirkungen auf das ethische Verhalten von Managern nehmen. Dies könnte dann der Fall sein, wenn der Code Schwächen aufweist und ein Manager diese Schwächen absichtlich für eigene Zwecke missbraucht. Auch das Verfolgen von schlechten Codes und damit verbundenes unbeabsichtigtes schlechtes Verhalten im Sinne des Unternehmens könnte eine Folge sein (vgl. Wotruba et al., 2001, S. 67).

4.4. Gruppenbasierte Belohnungen

Gruppenbasierte Belohnungen sollen die Zielkongruenz erhöhen und können so zu einer wahrheitsgemäßen Berichterstattung beitragen. Eine Falschberichterstattung im Rahmen von gruppenbasierten Belohnungen beeinflusst nicht nur die Wohlfahrt des einzelnen Managers, sondern die Wohlfahrt aller, die an der Entlohnung partizipieren. Sozialer Druck soll somit die Manager dazu motivieren, ihr Misreporting einzuschränken (vgl. Boster et al., 2018, S. 697 f.). So sind Manager zwar an der Maximierung ihres eigenen Nutzens interessiert, erleiden gleichzeitig aber auch einen Nutzenverlust, wenn ihre Unehrlichkeit den Ertrag anderer negativ beeinflusst (vgl. Gneezy, 2005, S. 391). Auch Maas und Van Rinsum (2013) erkennen dies als einen Faktor, der Misreporting verringern kann. So überschätzen Manager ihre Performance weniger stark, wenn dies den Pay-Off ihrer Kollegen verringern würde (vgl. Maas & Van Rinsum, 2013, S. 1179).

Bei gruppenbasierten Belohnungen ergibt sich die Belohnung außerdem aus den Entscheidungen aller Beteiligten. Manager haben deshalb einen Anreiz dazu gegenseitig Informationen auszutauschen (vgl. Kelly, 2010, S. 46). Zusammenarbeit wird im Unternehmenskontext ohnehin immer wichtiger, da sich das Unternehmen erhofft, dadurch Synergien zu heben. Entscheidend für eine gute Zusammenarbeit ist Vertrauen zwischen den einzelnen Parteien (vgl. Coletti, Sedatole & Towry, 2005, S. 478). Der Einsatz von Control-Instrumenten führt zu einem gesteigerten Kooperationsbedarf zwischen den beteiligten Parteien, was wiederum das Vertrauen fördert (vgl. ebda., S. 496). Gruppenbasierte Entlohnungen können so durch zugrunde liegendes Teamwork zwischen den Managern helfen, dieses Vertrauen herzustellen (vgl. Merchant & Van der Stede, 2017, S. 99).

Eine Problematik bei der Arbeit in Teams stellt für Unternehmen die nicht immer eindeutige Zuordnung von Arbeitsleistung zu Arbeitsergebnis dar (vgl. Rajan & Reichelstein, 2006, S. 603). Dies kann ein Problem bei der individuellen Anreizgestaltung darstellen. Eine Alternative stellt dann ein Anreizsystem auf Gruppenebene dar (vgl. Towry, 2003, S. 1070). Belohnungen auf Gruppenebene sind aus Unternehmenssicht folglich wünschenswert, da es die Berechnung individueller Performanceanteile und -beurteilungen erübrigt. Zugleich hat dies den Vorteil, dass einzelne Manager dadurch keinen Anreiz haben ihre Performance auf Kosten der gesamten Unternehmensperformance zu erhöhen. Ein Nachteil von gruppenbasierten Entlohnungen liegt in einer Free-Rider-Problematik, wodurch sich einzelne Manager negativ auf das Ergebnis der Gruppe auswirken, gleichzeitig aber vom Aufwand der Gruppe profitieren (vgl. Fisher, Peffer & Sprinkle, 2003, S. 52).

Auch einer weiteren Gefahr sollten sich Unternehmen bei der Anwendung von gruppenbasierten Entlohnungen bewusst sein: Wenn es einem Manager gelingt, sein unethisches Verhalten vor sich selbst zu rechtfertigen, so kann dies sein Misreporting erhöhen. Eine Möglichkeit dazu besteht, wenn der berichtete Slack zwischen mehreren Personen aufgeteilt wird. Dabei ist es unerheblich, ob die anderen Mitarbeiter Wissen über das Misreporting des Managers haben oder nicht. Der Manager kann in diesem Fall sein Misreporting mit dem gemeinsamen Interesse aller rechtfertigen. Beeinflusst wird die Entscheidung zum Misreporting des Managers aber auch von den Charaktereigenschaften der anderen Mitarbeiter. Sind diese dem Manager als besonders ehrlich bekannt, so wird auch der Manager zu einer ehrlicheren Berichterstattung tendieren. Dies kann damit Ansatzpunkt für Unternehmen darstellen, um ihre Manager zu einer wahrheitsgemäßen Berichterstattung zu motivieren (vgl. Church, Hannan & Kuang, 2012, S. 164 f.).

Weitere Forschung erscheint sinnvoll, um zu untersuchen, wie sich gruppenbasierte Entlohnungen auf eine wahrheitsgemäße Berichterstattung auswirken (vgl. Feichter & Grabner, 2020, S. 163), Bisherige Ergebnisse legen aber nahe, dass Instrumente wie Profit-Sharing Pläne einen Beitrag zu einem geringeren Maß an Misreporting leisten können (vgl. Evans et al., 2001, S. 554; Boster et al., 2018, S. 712).

4.5. Arbeitsplatzgestaltung und Kommunikation von Berichten

Soziale Normen kommen auch in der Gestaltung der Arbeitsumgebung in Unternehmen zum Tragen. Dabei richten sich soziale Normen nach dem Verhalten und den Wertvorstellungen des Großteils der Peer-Group. Ein Manager wird entsprechend versuchen, diese Normen zu erfüllen. Die sozialen Normen unterscheiden sich dabei je nach Kultur, Gesellschaft, etc. Je nach Kulturkreis sollten Unternehmen berücksichtigen, dass auch eine soziale Norm zur Unehrlichkeit vorherrschen kann (vgl. Altenburger, 2017, S. 9 f.). In einer offenen Arbeitsplatzgestaltung besteht die Möglichkeit, dass Mitarbeiter Teile der Kommunikation ihrer Kollegen mitverfolgen. Entsprechend besteht die Möglichkeit, dass ein Agent von den Berichten eines anderen Agenten an die Zentrale erfährt. Eine offene Arbeitsplatzgestaltung kann dabei viele Formen einnehmen, z. B. die Anzahl der Manager, die in einen E-Mail-Verteiler mitaufgenommen werden, oder offene Großraumbüros (vgl. Evans, Moser, Newman & Stikeleather, 2016, S. 1376). Überlegungen zur Gestaltung des Arbeitsumfeldes kann ein Unternehmen im Sinne der Cultural Controls in die Ausgestaltung seines MCS einfließen lassen. Da dadurch ein Mutual-Monitoring-Effekt eintreten kann, sind sie eng mit den in Kapitel 4.4 behandelten gruppenbasierten Belohnungen verknüpft (vgl. Merchant & Van der Stede, 2017, S. 99). In einer Studie von Maas und Van Rinsum (2013) zeigte sich, dass Manager nicht den maximal möglichen Betrag an Slack abgreifen. Als einen Faktor dafür machen die Autoren

die Offenheit der Berichtsumgebung aus. So neigten Manager weniger zur Überschätzung ihrer Performance, wenn die abgegebenen Berichte von den anderen Managern eingesehen werden können. Dies deutet darauf hin, dass Manager nicht nur ihren eigenen Nutzen maximieren wollen, sondern auch einen Disnutzen davontragen, wenn sie soziale Normen verletzen oder sich in ihrer Arbeitsumgebung unfair verhalten (vgl. Maas & Van Rinsum, 2013, S. 1179). Ma (1988) hat ein Modell aufgestellt, in welchem Agenten ein finanzieller Anreiz geboten wird, das Misreporting der anderen Agenten aufzudecken. In einer Berichtsumgebung, in der Agenten die Berichte dar anderen Agenten einsehen können, müsste demnach der ökonomischen Nutzenoptimierer-Theorie zufolge, ein Agent beobachtetes Fehlverhalten stets melden, da er dadurch seinen Nutzen maximiert. Dieses Vorgehen würde dazu führen, dass ein Agent seinerseits immer wahrheitsgemäß berichtet, da sein Verhalten anderenfalls stets aufgedeckt werden würde (vgl. Ma, 1988, S. 599 f.).

Eine Reihe von Studien beschäftigt sich mit Informationsoffenheit sowie anderen Faktoren im Unternehmen und deren Einfluss auf die wahrheitsgemäße Berichterstattung.

Evans et al. (2016) untersuchen eine Situation, in der zwei Manager ihre Kosten sowie die Kosten des jeweiligen anderen Managers kennen und diese an die Zentrale melden sollen. Grundsätzlich wird erwartet, dass in dieser Situation wahrheitsgemäße Berichterstattung für beide Manager die dominante Strategie darstellt. Die Autoren variieren in ihrer Untersuchung die Offenheit der Arbeitsplatzgestaltung in der Form, dass die Manager die Berichte des anderen Managers an die Zentrale entweder einsehen können oder nicht (vgl. Evans et al., 2016, S. 1376 f.). Dabei zeigt sich, dass eine offene Arbeitsplatzgestaltung aufgrund der stattfindenden Kommunikation zwischen den Agenten zu Verlusten für das Unternehmen führt (vgl. Evans et al., 2016, S. 1397).

Eine ähnliche Untersuchung stellen Paz et al. (2014) an. Sie unterscheiden zwischen keiner Veröffentlichung von Berichten anderer Manager sowie der vollständigen und der teilweisen Veröffentlichung der Berichte. Als teilweise Veröffentlichung von Berichten wird lediglich die Veröffentlichung der Berichte der am wenigsten ehrlichen Manager bezeichnet (vgl. Paz et al., 2014, S. 86). Auch in dieser Studie fanden die Autoren keine Belege dafür, dass die Veröffentlichung von Berichten die Ehrlichkeit im Unternehmen steigert. Stattdessen führte dies zu einer Verringerung der Ehrlichkeit (vgl. Paz et al., 2014, S. 110). Wenn sowohl die Berichte der ehrlichen als auch unehrlichen Manager veröffentlicht werden, würden sich Manager an den unehrlichen Berichten orientieren und diese dazu nutzen ihr eigenes unethisches Verhalten zu rechtfertigen. Die Autoren bezeichnen dies auch als eine Umkehrung der sozialen Norm. Die Veröffentlichung von Berichten und damit die Möglichkeit das eigene Reporting-Verhalten mit dem von unethischen Agenten zu vergleichen, senkt die Kosten des Agenten sich selbst unethisch zu verhalten. Manager überdenken dann ihre eigenen Vorstellungen davon, was ethisches Verhalten ist, was in einem verstärkten unethischen Verhalten resultiert (vgl. Paz et al., 2014, S. 110 f.). Brunner und Ostermaier (2019b) ergänzen, dass Manager bei einer teilweisen Veröffentlichung der Berichte ihrer Kollegen unehrlicher berichten würden als bei voller Transparenz. Die Agenten würden demnach fehlende Informationen mit eigenen Annahmen füllen und den fehlenden Berichten Misreporting unterstellen. Manager erwarten von ihren Kollegen im Unternehmen entsprechend keine totale Ehrlichkeit, sondern rechnen bereits mit einem gewissen Anteil von Misreporting (vgl. Brunner & Ostermaier, 2019a, S. 142).

Schwering (2017) untersucht, ob sich die Ehrlichkeit in den verschiedenen Situationen im Laufe der Zeit verändert. Sie stellt fest, dass sich die Ehrlichkeit bei keiner Veröffentlichung der Berichte nicht verändert, bei anonymer sowie nicht anonymer Veröffentlichung sinkt die Ehrlichkeit im Zeitablauf. Als Begründung für das Absinken der Ehrlichkeit ließen sich sowohl finanzielle Anreizfaktoren als auch Druck durch soziale Normen finden (vgl. Schwering, 2017, S. 1169 f.).

In Zhang (2008) Experiment bestätigt sich, dass Agenten ihrerseits nicht immer unehrliches Verhalten ihrer Mitagenten an den Prinzipal melden, selbst wenn dies aus rein finanzieller Ansicht den größten Anreiz darstellen würde (vgl. Zhang, 2008, S. 1143).

Die aufgeführten Studien haben gemeinsam, dass sich, anders als erwartet, ein Anstieg an Misreporting durch Offenlegung der Berichte an andere Manager zeigt. Unternehmen sollten entsprechend berücksichtigen, dass neben den positiven Effekten wie Koordination und Vertrauen zwischen Managern, auch potenzielle Kosten entstehen. Setzt das Unternehmen auf ein Peer-Monitoring-Control-System so kann eine zu offene Arbeitsplatzgestaltung die Vorteile dieses Systems zunichtemachen (vgl. Evans et al., 2016, S. 1397). Außerdem motiviert die Offenlegung von Berichten Manager dazu sich an „schlechten“ Beispielen zu orientieren, die die vom Unternehmen gewünschten Normen nicht vertreten, anstelle von „guten“ Beispielen, die ehrlich berichten (vgl. Brunner & Ostermaier, 2019a, S. 142). Unternehmen sollten deswegen unethisches Verhalten sehr genau im Auge behalten. Selbst wenn es sich um kleine Vergehen handelt (z. B. überhöhte Reisekosten) können diese Vergehen langfristig das Verhalten der anderen Manager im Unternehmen negativ beeinflussen (vgl. Paz et al., 2014, S. 111 f.). Aus den Ergebnissen von Schwering (2017) lässt sich ableiten, dass Agenten an ihrer Ehrlichkeitspräferenz festhalten, wenn sie keine Informationen über die Ehrlichkeit anderer Agenten besitzen. Bei einer Offenlegung sinkt die individuelle Ehrlichkeit hingegen. Werden bspw. Anreizsysteme an das Budget gekoppelt, sollte darauf geachtet werden, wie eng Mitarbeiter zusammenarbeiten, v. a. wie deren Kommunikation untereinander erfolgt, und ob ein Informationsaustausch bezüglich des Berichtsverhaltens stattfindet. (vgl. Schwering, 2017, S. 1170).

Maas und Van Rinsum (2013) erwarten, dass sich eine offene Informationspolitik für ein Unternehmen als vorteilhaft erweisen könnte, da dies den Druck zu wahrheitsgemäßer Berichterstattung erhöht. Wird Managern eine intrinsische Motivation zur Ehrlichkeit unterstellt, stelle dies eine Art Selbstregulierung dar, da der Manager dann nicht erwi-

scht werden wolle (vgl. Maas & Van Rinsum, 2013, S. 1180). Doch genau diese Ehrlichkeitspräferenz wird bei einer offenen Arbeitsplatzgestaltung durch Gruppenzwang und die Umkehrung sozialer Normen herausgefordert, was in verstärktem Misreporting resultiert (vgl. Schwering, 2017, S. 1169 f.). Außerdem kann die Wahrnehmung von Fairness im Unternehmen entscheidend beeinflussen, ob und in welchem Ausmaß sich Mitarbeiter unethisch verhalten (vgl. Cohen, Holder-Webb, Sharp & Pant, 2007, S. 1130). Mit Transparenz sollte entsprechend vorsichtig umgegangen werden, da diese unethische Tendenzen verstärken kann (vgl. Brunner & Ostermaier, 2019a, S. 142). Ein Unternehmen profitiert in einem Peer-Monitoring System folglich dann, wenn es ihm gelingt die Anreize zu opportunistischem Verhalten der einzelnen Manager möglichst gering zu halten. Die Motivierung von Agenten beobachtetes Missverhalten aufzudecken und zu melden, stellt hingegen kein optimales Control-Instrument dar (vgl. Zhang, 2008, S. 1144).

Möglichkeiten zur weiteren Forschung finden sich in den Ausgestaltungen der Peer-Monitoring-Systeme. So könnte es eine mögliche Alternative darstellen, lediglich die Berichte der ehrlichsten Manager zu veröffentlichen (vgl. Paz et al., 2014, S. 112). Denkbar wäre, dass dadurch eine soziale Norm aktiviert wird und die Agenten infolgedessen ehrlicher berichten.

Ein weiterer Punkt, der mit der Beobachtbarkeit von Informationen einhergeht, stellen Gehaltsunterschiede zwischen Managern dar. Guo, Libby und Liu (2017) untersuchen, ob das wissentliche Vorliegen von Gehaltsunterschieden einen Einfluss auf die wahrheitsgemäße Berichterstattung nimmt. Dabei unterscheiden sie zwischen hohen und niedrigen Gehaltsunterschieden. Ist der Unterschied groß und dem betrachteten Manager ist dies bewusst, so enthält dessen Bericht mehr Slack. Die Autoren führen ihre Beobachtungen auf das Fairnessempfinden der Manager zurück. Unternehmen sollten beim Design ihrer Anreizsysteme auch ihre interne Reportingumgebung bedenken. Bei hohen Gehaltsunterschieden unter den Managern hätte demnach eine offene Informationspolitik von Seiten des Unternehmens einen negativen Einfluss auf die Ehrlichkeit der Manager, während bei geringen Gehaltsunterschieden eine offene Informationspolitik einen positiven Einfluss auf die Ehrlichkeit der Manager nehmen kann (vgl. Guo et al., 2017, S. 597).

4.6. Tone at the Top

Auch durch eine passende Ausgestaltung des Tone at the Top im Unternehmen kann das Verhalten der Mitarbeiter beeinflusst werden (vgl. Merchant & Van der Stede, 2017, S. 100). Bei einem gut ausgestalteten Tone at the Top wird das Top-Management als Vertreter des Unternehmens akzeptiert und erfährt dadurch eine Legitimität für sein Handeln. Diese führt dazu, dass Mitarbeiter akzeptieren, dass das Top-Management an den erzielten Gewinnen partizipiert (vgl. Douthit & Majerczyk, 2019, S. 11). Die Unternehmenskultur wird u. a. durch das Top Management geprägt. Mitarbeiter beobachten das Verhalten des Top-Managements, was deren Vorstellung von der Unternehmenskultur beeinflusst. Deshalb sollte das Top Management stets die Werte verkörpern, die es im Unternehmen implementieren will (vgl. Zablow, 2006, S. 26). Andernfalls kann dies das unethische Verhalten unter den Mitarbeitern erhöhen, diese suchen dann nach eigenen Möglichkeiten sich unethisch zu verhalten (vgl. Greenfield, Norman & Wier, 2008, S. 422).

Einen wichtigen Bereich für die Unternehmenssteuerung stellt die Kommunikation zwischen dem Top-Management und den Managern dar. Bereits bevor Manager einen Bericht an das Top-Management abgeben, kann dieses durch eine geschickte Kommunikation zu einer wahrheitsgemäßen Berichterstattung beitragen. Antizipiert das Top-Management bspw. vor Abgabe eines Budgets, dass der Manager plant, Slack in sein Budget einzubauen, so kann es durch eine entsprechende Kommunikation den finalen Bericht des Managers noch beeinflussen. Wichtig ist, dass das Top-Management dazu vor Abgabe des Budgets ein entsprechendes Signal vom Agenten empfangen muss. Findet keinerlei Informationsaustausch statt, kann das Top-Management auch keine Strömungen erkennen und somit auch keinen Einfluss mehr auf das Verhalten der Mitarbeiter ausüben (vgl. Kachelmeier, Smith & Yancey, 1994, S. 167 f.). Auch Rankin, Schwartz und Young (2003) unterstützen diese These. Demnach könnte im Rahmen der Budgetierung Kommunikation zwischen Budgetverantwortlichem und Vorgesetztem als ein informelles Control-Instrument eine wichtige unterstützende Rolle spielen (vgl. Rankin et al., 2003, S. 87-89).

Im Zuge der Kommunikation können Anweisungen von Vorgesetzten unethisches Verhalten auch begünstigen. Manager berichten weniger ehrlich, wenn ihnen dies von einer höheren Ebene beauftragt worden ist. Sie können dann die Verantwortung dafür von sich weisen und besitzen einen Rechtfertigungsgrund dafür. Schlechte Vorbilder und eine schlechte Kommunikation von Seiten des Top-Managements sollten deshalb vermieden werden, um die Tendenz zum Misreporting nicht zu verstärken (vgl. Mayhew & Murphy, 2014, S. 439).

Durch ihren Einfluss auf die Unternehmenskultur kann das Top-Management die Wirkung der in Kapitel 4.3 besprochenen Codes of Conduct verstärken. So kann eine klare Kommunikation der Codes den Mitarbeitern verdeutlichen, dass ethisches Verhalten dem Top-Management ein Anliegen ist und dadurch das Verhalten der Mitarbeiter steuern (vgl. Adams et al., 2001, S. 208). Schwartz, Dunfee und Kline (2005) schlägt vor, dass Unternehmen einen Code of Conduct speziell für das Top-Management integrieren sollten, ähnlich dem Code für das ganze Unternehmen. Dies könnte dazu beitragen, im gesamten Unternehmen ein ethisches Arbeitsklima zu schaffen und damit auch Werte wie Ehrlichkeit zu fördern (vgl. Schwartz et al., 2005, S. 96).

Auch der relative Lohnunterschied des Top-Managements zu untergeordneten Ebenen kann eine Rolle bei der wahrheitsgemäßen Berichterstattung spielen. Eine Studie von Guo, Libby, Liu und Tian (2020) legt nahe, dass Manager bei einem großen relativen Lohnunterschied eher zu Misreporting neigen als bei geringen Unterschieden (vgl. Guo et al., 2020, S. 572 f.). Bei Betrachtung des absoluten Gehalts

zeigt sich, dass Agenten umso ehrlicher berichten, je höher ihr Gehalt ist. Zusätzlich führt ein hohes Gehalt dazu, dass Manager den Prinzipal als fairer einschätzen (vgl. Zhang, 2008, S. 1143).

Die Fairness des Top-Managements zu den Managern spielt entsprechend ebenfalls eine Rolle, wenn eine wahrheitsgemäße Berichterstattung erreicht werden soll. Verhaltenswissenschaftler argumentieren, dass die Bereitschaft von Agenten ihre eigenen Informationen sowie beobachtetes Verhalten von anderen Agenten wahrheitsgemäß zu berichten, damit zusammenhängt als wie fair sie den Prinzipal ansehen (vgl. Zhang, 2008, S. 1126). Im Rahmen von Whistleblowing-Aktivitäten könnte das Top-Management entsprechend als Ansprechpartner für ihre Mitarbeiter zur Verfügung stehen (vgl. Kaptein, 2011, S. 527). Empfinden Manager allerdings das Top-Management als unfair, erhöht dies die Chance, dass Manager Gelegenheiten nutzen sich opportunistisch zu verhalten (vgl. Nagin et al., 2002, S. 870). So könnten sich Manager im Unternehmen zusammenschließen und gegen den Prinzipal arbeiten, wenn sie sich von diesem ungerecht behandelt fühlen (vgl. Maas & Yin, 2021, S. 14 f.).

Eine darauf verstärkende Wirkung kann eintreten, wenn die Agenten miteinander kommunizieren können. Wenn eine Kommunikation stattfindet und der Prinzipal von den Agenten als unfair eingestuft wird, so kann dies den Grad an Unehrlichkeit im Unternehmen erhöhen (vgl. Zhang, 2008, S. 1126). Eine hohe Transparenz führt im Unternehmen grundsätzlich zu einem besseren Informationsaustausch zwischen den Managern. Eine steigende Transparenz mit gleichzeitig unfairer Behandlung der Mitarbeiter erhöht das Risiko von Absprachen zwischen den Managern zuungunsten des Unternehmens (vgl. Maas & Yin, 2021, S. 14). Für das Top-Management ergibt sich dadurch die Herausforderung, aus einer erhöhten Transparenz im Unternehmen keinem erhöhten Misreporting zu unterliegen. Dazu ist es wichtig, auf eine gute Behandlung der Mitarbeiter zu setzen (vgl. Maas & Yin, 2021, S. 15).

Nutzt das Unternehmen ein vertikales Anreizsystem, so sollen Manager dazu motiviert werden, beobachtetes Fehlverhalten an die Vorgesetzten zu melden (vgl. Towry, 2003, S. 1070). Nach der klassischen Agency-Theorie besteht für jeden Manager ein Anreiz sich ehrlich zu verhalten, da demnach jeder Manager stets befürchten muss, dass ein anderer Manager sein Fehlverhalten aufdeckt und der Zentrale meldet. In der Realität ist jedoch denkbar, dass Manager wie auch bereits in Kapitel 4.5 dargelegt aufgrund guter Beziehungen zueinander, füreinander einstehen und Fehlverhalten entsprechend nicht weiter melden (vgl. Towry, 2003, S. 1070). Fairnessbedenken könnten diesen Einfluss noch verstärken. Ein vertikales Anreizsystem wäre dadurch weniger effektiv und könnte sogar in einem erhöhten Misreporting resultieren (vgl. Towry, 2003, S. 1088).

5. Fazit und Ausblick

Um ihre Ziele erreichen zu können, sind Unternehmen auf verlässliche Informationen angewiesen. Dazu gehört aufgrund von Informationsasymmetrien auch die wahrheitsgemäße Weitergabe privater Informationen von Mitarbeitern und anderen Managern. In der experimentellen, verhaltenswissenschaftlichen Forschung finden sich zahlreiche Hinweise darauf, dass Manager Ehrlichkeitspräferenzen aufweisen. In dieser Arbeit wurde untersucht, wie die Unternehmensführung durch eine geschickte Ausgestaltung von Steuerungsinstrumenten zu einer wahrheitsgemäßen Berichterstattung beitragen kann.

Entgegen den Annahmen der Prinzipal-Agenten-Theorie hat sich in der verhaltenswissenschaftlichen Forschung gezeigt, dass Manager Ehrlichkeitspräferenzen aufweisen. Es handelt sich bei Managern also nicht um rein eigennutzinteressierte Individuen, die bei jeder sich bietenden Gelegenheit lügen, wenn sich ihnen daraus nur ein Vorteil bietet. Vielmehr erleiden sie zunächst einen Disnutzen, wenn die Ehrlichkeitspräferenz unterdrückt wird. Ein Problem für das Unternehmen im Rahmen der Budgetierung stellt der Budgetary Slack dar. Dieser beschreibt die Situation, dass Manager ihre Zielvorgaben bewusst niedriger (bzw. im Falle von Kostenbudgets höher) verhandeln, als es ihnen mit den vorhandenen Ressourcen möglich wäre (vgl. Onsi, 1973, S. 535). Unternehmen wollen Slack so gering wie möglich halten. Steuerungsinstrumente, deren Gesamtheit im MCS eines Unternehmens abgebildet werden, zielen darauf ab, das Verhalten der Mitarbeiter im Sinne des Unternehmens zu beeinflussen. Sie können dem Unternehmen demzufolge dabei helfen Slack zu verringern und zu einer wahrheitsgemäßen Berichterstattung beizutragen.

Im Rahmen von formellen Controls ist es den Unternehmen möglich, das Verhalten ihrer Mitarbeiter über Results Controls und Action Controls zu steuern. Results Controls stellen Steuerungsinstrumente dar, bei denen Mitarbeiter für das Erreichen bestimmter Ziele belohnt werden. Gerade auf das Setzen von Zielvorgaben nehmen Mitarbeiter durch die von ihnen abgegebenen Berichte einen großen Einfluss. Um ihre Manager zu einer wahrheitsgemäßen Weitergabe der privaten Informationen zu motivieren, sollten Unternehmen hierbei auf eine faire Vertragsgestaltung achten und die Verträge auf die individuellen Ehrlichkeitspräferenzen ihrer Manager anpassen. Somit kann verhindert werden, dass (finanzielle) Anreize die Tendenz zur Ehrlichkeit untergraben.

Action Controls zielen konkret darauf ab, wie bestimmte Handlungen durchgeführt werden sollten. Eine direkte Einflussnahme auf Aktionen und Beschränkungen der Manager erscheint aufgrund der komplexen Tätigkeiten sowie des bestehenden Informationsvorsprungs wenig sinnvoll. Hingegen stellen Pre-Action Reviews im Budgetprozess ein wirksames Steuerungsinstrument dar. Ein zu starkes Eingreifen in den Aktionsraum des Agenten sollte hier jedoch auch nicht erfolgen, um den Manager nicht zu demotivieren. Ebenso kann die Zuweisung von Verantwortlichkeiten an Mitarbeiter und Kontrolle durch Audits zu einer wahrheitsgemäßen Bericht-

erstattung beitragen. Audits sollten dabei am besten von unternehmensexternen Parteien durchgeführt werden, für die Zuweisung von Verantwortlichkeiten erscheinen klare Definitionen und eine angemessene Kommunikation als sinnvoll, um Konflikten vorzubeugen.

Die Wirkungsweise von informellen Controls ergibt sich durch soziale Normen und sozialen Druck. Um eine wahrheitsgemäße Berichterstattung sicher zu stellen, sollte das Unternehmen in Form von Selektion und Weiterbildung ihrer Mitarbeiter für die Existenz ethisch geschulter Mitarbeiter im Unternehmen sorgen, wodurch ein Self-Monitoring-Effekt eintritt. Bezeichnet werden diese Steuerungsinstrumente als Personnel Controls.

Cultural Controls auf der anderen Seite führen zu einem Mutual-Monitoring-Effekt. Die Unternehmensführung besitzt hier diverse Ansatzpunkte. Die Implementierung und Kommunikation von Codes of Conduct kann zu einem ethischeren Arbeitsklima beitragen. Auch gruppenbasierte Entlohnungen können durch die Wirkung von sozialen Normen zu einer wahrheitsgemäßen Berichterstattung führen. Bei einer offenen Arbeitsplatzgestaltung sollte das Unternehmen auf die Einsehbarkeit der Berichte und die Kommunikation der Manager untereinander achten, da dies zu Misreporting führen kann. Zuletzt kann auch ein gut ausgestalteter Tone at the Top eine wahrheitsgemäße Berichterstattung unterstützen, v. a. in Kombination mit anderen Steuerungsinstrumenten.

Insgesamt betrachtet zeigt sich, dass Manager Ehrlichkeitspräferenzen aufweisen und sich auch fair verhalten wollen. Um davon zu profitieren, sollte das Unternehmen sich selbst fair verhalten und Maßnahmen umsetzen, um die sozialen Normen im Unternehmen zu verbessern. Passend dazu lässt sich festhalten, dass Steuerungsinstrumente nicht einzeln betrachtet werden sollten, da sie zueinander in Beziehungen stehen. Eine ganzheitliche Betrachtung des MCS eines Unternehmens schafft dementsprechend ein Verständnis dafür, wie Control-Instrumente auf das gesamte Unternehmen wirken.

Ehrlichkeitspräferenzen sind kontextabhängig und für die Forschung ist es Aufgabe und Herausforderung zugleich, diese Faktoren herauszufinden und zu evaluieren. Weitere Forschung könnte hier helfen, zu einer effektiveren Vertragsgestaltung beizutragen, die sich stärker auf die intrinsische Motivation von Individuen stützt (vgl. Hannan et al., 2006, S. 912). Gerade in der wirtschaftswissenschaftlichen Forschung stützen sich Forschungen häufig auf das Prinzipal-Agenten-Modell. Die Nutzung weiterer theoretischer Rahmenkonzepte könnte hier weitere Erkenntnisse liefern (vgl. Covaleski et al., 2003, S. 7).

Literatur

Abernethy, M. A., Bouwens, J. & Van Lent, L. (2004). Determinants of control system design in divisionalized firms. *The Accounting Review*, *79* (3), 545–570.

Abernethy, M. A. & Brownell, P. (1997). Management control systems in research and development organizations: The role of accounting, behavior and personnel controls. *Accounting, Organizations and Society*, *22* (3-4), 233–248.

Abernethy, M. A., Dekker, H. C. & Schulz, A. K.-D. (2015). Are employee selection and incentive contracts complements or substitutes? *Journal of Accounting Research*, *53* (4), 633–668.

Adams, J. S., Tashchian, A. & Shore, T. H. (2001). Codes of ethics as signals for ethical behavior. *Journal of Business Ethics*, *29* (3), 199–211.

Altenburger, M. (2017). The effect of injunctive social norms and dissent on budget reporting honesty. *Journal of International Accounting Research*, *16* (2), 9–31.

Antle, R. & Fellingham, J. (1995). Information rents and preferences among information systems in a model of resource allocation. *Journal of Accounting Research*, *33*, 41–58.

Arnold, M. C. & Schreiber, D. (2013). Audits, reputation, and repeated interaction in a capital budgeting setting. *European Accounting Review*, *22* (1), 185–213.

Baiman, S. (1990). Agency research in managerial accounting: A second look. *Accounting, Organizations and Society*, *15* (4), 341–371.

Baiman, S. & Lewis, B. L. (1989). An experiment testing the behavioral equivalence of strategically equivalent employment contracts. *Journal of Accounting Research*, *27* (1), 1–20.

Berry, C. M., Sackett, P. R. & Wiemann, S. (2007). A review of recent developments in integrity test research. *Personnel Psychology*, *60* (2), 271–301.

Blay, A., Douthit, J. & Fulmer, B. (2019). Why don't people lie? Negative affect intensity and preferences for honesty in budgetary reporting. *Management Accounting Research*, *42*, 56–65.

Blum, U., Dudley, L., Leibbrand, F. & Weiske, A. (2005). *Angewandte Institutionenökonomik*. Wiesbaden, Gabler.

Boster, C., Majerczyk, M. & Tian, Y. (2018). The effect of individual and pooled profit sharing plans on honesty in managerial reporting. *Contemporary Accounting Research*, *35* (2), 696–715.

Brink, A. G., Coats, J. C. & Rankin, F. W. (2018). Who's the boss? The economic and behavioral implications of various characterizations of the superior in participative budgeting research. *Journal of Accounting Literature*, *41*, 89–105.

Brown, J. L., Evans, J. H. & Moser, D. V. (2009). Agency theory and participative budgeting experiments. *Journal of Management Accounting Research*, *21* (1), 317–345.

Brown, J. L., Fisher, J. G., Peffer, S. A. & Sprinkle, G. B. (2017). The effect of budget framing and budget-setting process on managerial reporting. *Journal of Management Accounting Research*, *29* (1), 31–44.

Brown, J. L., Fisher, J. G., Sooy, M. & Sprinkle, G. B. (2014). The effect of rankings on honesty in budget reporting. *Accounting, Organizations and Society*, *39* (4), 237–246.

Brüggen, A. & Luft, J. (2011). Capital rationing, competition, and misrepresentation in budget forecasts. *Accounting, Organizations and Society*, *36* (7), 399–411.

Brunner, M. & Ostermaier, A. (2019a). Peer influence on managerial honesty: The role of transparency and expectations. *Journal of Business Ethics*, *154* (1), 127–145.

Brunner, M. & Ostermaier, A. (2019b). Sabotage in capital budgeting: The effects of control and honesty on investment decisions. *European Accounting Review*, *28* (1), 71–100.

Cardinaels, E. (2016). Earnings benchmarks, information systems, and their impact on the degree of honesty in managerial reporting. *Accounting, Organizations and Society*, *52*, 50–62.

Cardinaels, E. & Jia, Y. (2016). How audits moderate the effects of incentives and peer behavior on misreporting. *European Accounting Review*, *25* (1), 183–204.

Cardinaels, E. & Yin, H. (2015). Think twice before going for incentives: Social norms and the principal's decision on compensation contracts. *Journal of Accounting Research*, *53* (5), 985–1015.

Chenhall, R. H. (2003). Management control systems design within its organizational context: findings from contingency-based research and directions for the future. *Accounting, Organizations and Society*, *28* (2-3), 127–168.

Chong, V. K. & Wang, I. Z. (2019). Delegation of decision rights and misreporting: The roles of incentive-based compensation schemes and responsibility rationalization. *European Accounting Review*, *28* (2), 275–307.

Chonko, L. B., Wotruba, T. R. & Loe, T. W. (2003). Ethics code familiarity and usefulness: Views on idealist and relativist managers under varying conditions of turbulence. *Journal of Business Ethics*, *42* (3), 237–252.

Chow, C. W., Cooper, J. C. & Haddad, K. (1991). The effects of pay schemes and ratchets on budgetary slack and performance: A multiperiod experiment. *Accounting, Organizations and Society*, *16* (1), 47–60.

Chow, C. W., Cooper, J. C. & Waller, W. S. (1988). Participative budgeting: Effects of a truth-inducing pay scheme and information asymmetry on slack and performance. *The Accounting Review*, *63* (1), 111–122.

Christ, M. H., Emett, S. A., Summers, S. L. & Wood, D. A. (2012). The effects of preventive and detective controls on employee performance and motivation. *Contemporary Accounting Research*, *29* (2), 432–452.

Church, B. K., Hannan, R. L. & Kuang, X. J. (2012). Shared interest and honesty in budget reporting. *Accounting, Organizations and Society*, *37* (3), 155–167.

Church, B. K., Kuang, X. J. & Liu, Y. S. (2019). The effects of measurement basis and slack benefits on honesty in budget reporting. *Accounting, Organizations and Society*, *72*, 74–84.

Church, B. K., Lynn Hannan, R. & Kuang, X. (2014). Information acquisition and opportunistic behavior in managerial reporting. *Contemporary Accounting Research*, *31* (2), 398–419.

Cleek, M. A. & Leonard, S. L. (1998). Can corporate codes of ethics influence behavior? *Journal of Business Ethics*, *17* (6), 619–630.

Cohen, J. R., Holder-Webb, L., Sharp, D. J. & Pant, L. W. (2007). The effects of perceived fairness on opportunistic behavior. *Contemporary Accounting Research*, *24* (4), 1119–1138.

Coletti, A. L., Sedatole, K. L. & Towry, K. L. (2005). The effect of control systems on trust and cooperation in collaborative environments. *The Accounting Review*, *80* (2), 477–500.

Covaleski, M. A., Evans III, J. H., Luft, J. L. & Shields, M. D. (2003). Budgeting research: Three theoretical perspectives and criteria for selective integration. *Journal of Management Accounting Research*, *15* (1), 3–49.

Daumoser, C., Hirsch, B. & Sohn, M. (2018). Honesty in budgeting: A review of morality and control aspects in the budgetary slack literature. *Journal of Management Control*, *29* (2), 115–159.

Davidson, B. I. (2019). The effects of reciprocity and self-awareness on honesty in managerial reporting: Social value orientation matters. *Journal of Management Accounting Research*, *31* (1), 85–103.

Davila, T. & Wouters, M. (2005). Managing budget emphasis through the explicit design of conditional budgetary slack. *Accounting, Organizations and Society*, *30* (7-8), 587–608.

Dees, J. (1992). Ethics and agency theory: An introduction. In N. E. Bowie & R. Freeman (Hrsg.), *Principals, agents, and ethics* (S. 25–58). New York, Oxford University Press.

Doran, G. T. (1981). There's a S.M.A.R.T. way to write management's goals and objectives. *Management Review*, *70* (11), 35–36.

Douthit, J. D. & Majerczyk, M. (2019). Subordinate perceptions of the superior and agency costs: Theory and evidence. *Accounting, Organizations and Society*, *78*, 1–13.

Douthit, J. D. & Stevens, D. E. (2015). The robustness of honesty effects on budget proposals when the superior has rejection authority. *The Accounting Review*, *90* (2), 467–493.

Drake, A. R., Matuszewski, L. J. & Miller, F. (2014). The effect of personality traits and fairness on honesty in managerial reporting. *Advances in Management Accounting*, *22*, 43–69.

Dunk, A. S. & Perera, H. (1997). The incidence of budgetary slack: A field study exploration. *Accounting, Auditing & Accountability Journal*, *10* (5), 649–664.

Ebers, M. & Gotsch, W. (2019). Institutionenökonomische Theorien der Organisation. In A. Kieser & M. Ebers (Hrsg.), *Organisationstheorien* (S. 196–257). Stuttgart, W. Kohlhammer Verlag.

Evans, J. H., Hannan, R. L., Krishnan, R. & Moser, D. V. (2001). Honesty in managerial reporting. *The Accounting Review*, *76* (4), 537–559.

Evans, J. H., Moser, D. V., Newman, A. H. & Stikeleather, B. R. (2016). Honor

among thieves: Open internal reporting and managerial collusion. *Contemporary Accounting Research*, *33* (4), 1375–1402.

Ewert, R. & Wagenhofer, A. (2014). *Interne Unternehmensrechnung* (8. Aufl.). Berlin, Springer Gabler.

Falk, A. & Kosfeld, M. (2006). The hidden costs of control. *American Economic Review*, *96* (5), 1611–1630.

Feichter, C. & Grabner, I. (2020). Empirische Forschung zu Management Control–Ein Überblick und neue Trends. *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung*, *72* (2), 149–181.

Fiolleau, K., Libby, T. & Thorne, L. (2018). Dysfunctional behavior in organizations: Insights from the management control literature. *Auditing: A Journal of Practice & Theory*, *37* (4), 117–141.

Fisher, J. G. (1998). Contingency theory, management control systems and firm outcomes: Past results and future directions. *Behavioral Research in Accounting*, *10*, 47–64.

Fisher, J. G., Frederickson, J. R. & Peffer, S. A. (2006). Budget negotiations in multi-period settings. *Accounting, Organizations and Society*, *31* (6), 511–528.

Fisher, J. G., Maines, L. A., Peffer, S. A. & Sprinkle, G. B. (2002). Using budgets for performance evaluation: Effects of resource allocation and horizontal information asymmetry on budget proposals, budget slack, and performance. *The Accounting Review*, *77* (4), 847–865.

Fisher, J. G., Peffer, S. A. & Sprinkle, G. B. (2003). Budget-based contracts, budget levels, and group performance. *Journal of Management Accounting Research*, *15* (1), 51–74.

Fisher, J. G., Sprinkle, G. B. & Walker, L. L. (2007). Experimental budgeting research: Implications for practitioners. *Journal of Corporate Accounting & Finance*, *18* (6), 67–75.

Gallani, S., Krishnan, R., Marinich, E. J. & Shields, M. D. (2019). Budgeting, psychological contracts, and budgetary misreporting. *Management Science*, *65* (6), 2924–2945.

Garbers, Y. & Konradt, U. (2014). The effect of financial incentives on performance: A quantitative review of individual and team-based financial incentives. *Journal of Occupational and Organizational Psychology*, *87* (1), 102–137.

Gino, F., Krupka, E. L. & Weber, R. A. (2013). License to cheat: Voluntary regulation and ethical behavior. *Management Science*, *59* (10), 2187–2203.

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, *95* (1), 384–394.

Gneezy, U., Meier, S. & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives*, *25* (4), 191–210.

Greenfield, A., Norman, C. S. & Wier, B. (2008). The effect of ethical orientation and professional commitment on earnings management behavior. *Journal of Business Ethics*, *83* (3), 419–434.

Grover, S. L. (2005). The truth, the whole truth, and nothing but the truth: The causes and management of workplace lying. *Academy of Management Perspectives*, *19* (2), 148–157.

Gundlach, M. J., Douglas, S. C. & Martinko, M. J. (2003). The decision to blow the whistle: A social information processing framework. *Academy of Management Review*, *28* (1), 107–123.

Guo, L., Libby, T. & Liu, X. (2017). The effects of vertical pay dispersion: Experimental evidence in a budget setting. *Contemporary Accounting Research*, *34* (1), 555–576.

Guo, L., Libby, T., Liu, X. & Tian, Y. (2020). Vertical pay dispersion, peer observability, and misreporting in a participative budgeting setting. *Contemporary Accounting Research*, *37* (1), 575–602.

Haesebrouck, K. (2021). The effects of information acquisition effort, psychological ownership, and reporting context on opportunistic managerial reporting. *Contemporary Accounting Research*, *38* (4), 3085–3112.

Haesebrouck, K., Cools, M. & Van den Abbeele, A. (2018). Status differences and knowledge transfer: The effect of incentives. *The Accounting Review*, *93* (1), 213–234.

Hannan, R. L., Rankin, F. W. & Towry, K. L. (2006). The effect of information systems on honesty in managerial reporting: A behavioral perspective. *Contemporary Accounting Research*, *23* (4), 885–918.

Hannan, R. L., Rankin, F. W. & Towry, K. L. (2010). Flattening the organization: The effect of organizational reporting structure on budgeting effectiveness. *Review of Accounting Studies*, *15* (3), 503–536.

Helgesson, K. S. (2011). Public-private partners against crime: Governance, surveillance and the limits of corporate accountability. *Surveillance & Society*, *8* (4), 471–484.

Hobson, J. L., Mellon, M. J. & Stevens, D. E. (2011). Determinants of moral judgments regarding budgetary slack: An experimental examination of pay scheme and personal values. *Behavioral Research in Accounting*, *23* (1), 87–107.

Ittner, C. D. & Larcker, D. F. (2001). Assessing empirical research in managerial accounting: A value-based management perspective. *Journal of Accounting and Economics*, *32* (1-3), 349–410.

Jensen, M. C. (1983). Organization theory and methodology. *Accounting Review*, *58* (2), 319–339.

Jensen, M. C. (2003). Paying people to lie: The truth about the budgeting process. *European Financial Management*, *9* (3), 379–406.

Jensen, M. C. & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, *3* (4), 305-360.

Jollineau, S. J., Vance, T. W. & Webb, A. (2012). Subordinates as the first line of defense against biased financial reporting. *Journal of Management Accounting Research*, *24* (1), 1–24.

Kachelmeier, S. J., Smith, J. R. & Yancey, W. F. (1994). Budgets as a credible threat: An experimental study of cheap talk and forward induction. *Journal of Management Accounting Research*, *6*, 144–174.

Kahneman, D. & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47* (2), 263–291.

Kaptein, M. (2004). Business codes of multinational firms: What do they say? *Journal of Business Ethics*, *50* (1), 13–31.

Kaptein, M. (2011). From inaction to external whistleblowing: The influence of the ethical culture of organizations on employee responses to observed wrongdoing. *Journal of Business Ethics*, *98* (3), 513–530.

Keating, A. S. (1997). Determinants of divisional performance evaluation practices. *Journal of Accounting and Economics*, *24* (3), 243–273.

Kelly, K. (2010). The effects of incentives on information exchange and decision quality in groups. *Behavioral Research in Accounting*, *22* (1), 43–65.

Kilfoyle, E. & Richardson, A. J. (2011). Agency and structure in budgeting: Thesis, antithesis and synthesis. *Critical Perspectives on Accounting*, *22* (2), 183–199.

Koford, K. & Penno, M. (1992). Accounting, principal-agent theory, and self-interested behavior. In N. Bowie & R. Freeman (Hrsg.), *Ethics and agency theory: An introduction* (S. 127–142). New York, Oxford University Press.

Kreipl, C. (2020). *Verantwortungsvolle Unternehmensführung* (1. Aufl.). Wiesbaden, Springer Gabler.

Lange, D. (2008). A multidimensional conceptualization of organizational corruption control. *Academy of Management Review*, *33* (3), 710–729.

Lau, C. M. & Eggleton, I. R. (2003). The influence of information asymmetry and budget emphasis on the relationship between participation and slack. *Accounting and Business Research*, *33* (2), 91–104.

Liao, Y.-S. (2005). Business strategy and performance: The role of human resource management control. *Personnel Review*, *34* (3), 294–309.

Libby, T. (2015). The effect of fairness in contracting on the creation of budgetary slack. *Advances in Accounting Behavioral Research*, *6*, 145–169.

Libby, T. & Lindsay, R. M. (2010). Beyond budgeting or budgeting reconsidered? A survey of North-American budgeting practice. *Management Accounting Research*, *21* (1), 56–75.

Libby, T., Proell, C. A. & Smith, S. D. (2019). Difficult cost goals and conflicting honesty incentives: The influence of decision responsibility. *Journal of Management Accounting Research*, *31* (3), 185–200.

Liessem, T., Schedlinsky, I., Schwering, A. & Sommer, F. (2015). Budgetary slack under budget-based incentive schemes—the behavioral impact of social preferences, organizational justice, and moral disengagement. *Journal of Management Control*, *26* (1), 81–94.

Lohmann, C. (2013). Schemata zur wahrheitsgemäßen Berichterstattung und ihr Einfluss auf die Arbeitsleistung. *Wirtschaftswissenschaftliches Studium*, *42* (2), 56–62.

Lourenço, S. M. (2016). Monetary incentives, feedback, and recognition—Complements or substitutes? Evidence from a field experiment in a retail services company. *The Accounting Review*, *91* (1), 279–297.

Luft, J. L. (1997). Fairness, ethics and the effect of management accounting on transaction costs. *Journal of Management Accounting Research*, *9*, 199–216.

Ma, C.-T. (1988). Unique implementation of incentive contracts with many agents. *The Review of Economic Studies*, *55* (4), 555–572.

Maas, V. S. & Van Rinsum, M. (2013). How control system design influences performance misreporting. *Journal of Accounting Research*, *51* (5), 1159–1186.

Maas, V. S. & Yin, H. (2021). Finding partners in crime? How transparency about managers' behavior affects employee collusion. *Accounting, Organizations and Society*. (Online-Vorveröffentlichung)

Malmi, T. & Brown, D. A. (2008). Management control systems as a package—Opportunities, challenges and research directions. *Management Accounting Research*, *19* (4), 287–300.

Matějka, M., Merchant, K. A. & Van der Stede, W. A. (2009). Employment horizon and the choice of performance measures: Empirical evidence from annual bonus plans of loss-making entities. *Management Science*, *55* (6), 890–905.

Matuszewski, L. J. (2010). Honesty in managerial reporting: is it affected by perceptions of horizontal equity? *Journal of Management Accounting Research*, *22* (1), 233–250.

Mayhew, B. W. & Murphy, P. R. (2009). The impact of ethics education on reporting behavior. *Journal of Business Ethics*, *86* (3), 397–416.

Mayhew, B. W. & Murphy, P. R. (2014). The impact of authority on reporting behavior, rationalization and affect. *Contemporary Accounting Research*, *31* (2), 420–443.

Mazar, N., Amir, O. & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, *45* (6), 633–644.

Merchant, K. & Van der Stede, W. (2017). *Management Control Systems. Performance Measurement, Evaluation, and Incentives* (4. Aufl.). Harlow, Prentice Hall.

Milgrom, P. & Roberts, J. (1995). Complementarities and fit strategy, structure, and organizational change in manufacturing. *Journal of Accounting and Economics*, *19* (2-3), 179–208.

Mittendorf, B. (2006). Capital budgeting when managers value both honesty and perquisites. *Journal of Management Accounting Research*, *18* (1), 77–95.

Murphy, P. R., Wynes, M., Hahn, T.-A. & Devine, P. G. (2020). Why are people honest? Internal and external motivations to report honestly. *Contemporary Accounting Research*, *37* (2), 945–981.

Nagar, V. (2002). Delegation and incentive compensation. *The Accounting Review*, *77* (2), 379–395.

Nagin, D. S., Rebitzer, J. B., Sanders, S. & Taylor, L. J. (2002). Monitoring, motivation, and management: The determinants of opportunistic behavior in a field experiment. *American Economic Review*, *92* (4), 850–873.

Newman, A. H. (2014). An investigation of how the informal communication of firm preferences influences managerial honesty. *Accounting, Organizations and Society*, *39* (3), 195–207.

Nichol, J. E. (2019). The effects of contract framing on misconduct and entitlement. *The Accounting Review*, *94* (3), 329–344.

Onsi, M. (1973). Factor analysis of behavioral variables affecting budgetary slack. *The Accounting Review*, *48* (3), 535–548.

Otley, D. T. (1980). The contingency theory of management accounting: Achievement and prognosis. *Accounting, Organizations and Society*, *5* (4), 413–428.

Paul, J. (2014). *Beteiligungscontrolling und Konzerncontrolling* (1. Aufl.). Wiesbaden, Springer Gabler.

Paz, M., Reichert, B. E. & Woods, A. (2014). How does peer honesty affect focal manager honesty in a budget reporting setting? *Advances in Accounting Behavioral Research*, *16*, 85–114.

Perrow, C. (1970). *Organizational analysis: A sociological view*. Brooks/Cole.

Presslee, A., Vance, T. W. & Webb, R. A. (2013). The effects of reward type on employee goal setting, goal commitment, and performance. *The Accounting Review*, *88* (5), 1805–1831.

Rae, K. & Subramaniam, N. (2008). Quality of internal control procedures: Antecedents and moderating effect on organisational justice and employee fraud. *Managerial Auditing Journal*, *23* (2), 104–124.

Rajan, M. V. & Reichelstein, S. (2006). Subjective performance indicators and discretionary bonus pools. *Journal of Accounting Research*, *44* (3), 585–618.

Rankin, F. W., Schwartz, S. T. & Young, R. A. (2003). Management control using nonbinding budgetary announcements. *Journal of Management Accounting Research*, *15* (1), 75–93.

Rankin, F. W., Schwartz, S. T. & Young, R. A. (2008). The effect of honesty and superior authority on budget proposals. *The Accounting Review*, *83* (4), 1083–1099.

Salterio, S. & Webb, A. (2006). Honesty in Accounting and Control: A Discussion of "The Effect of Information Systems on Honesty in Managerial Reporting: A Behavioral Perspective". *Contemporary Accounting Research*, *23* (4), 919–932.

Sánchez-Expósito, M. J. & Naranjo-Gil, D. (2017). Effects of management control systems and cognitive orientation on misreporting: An experiment. *Management Decision*, *55* (3), 579–594.

Schäffer, U. (2013). Management accounting research in Germany: from splendid isolation to being part of the international community. *Journal of Management Control*, *23* (4), 291–309.

Schatzberg, J. W. & Stevens, D. E. (2008). Public and private forms of opportunism within the organization: A joint examination of budget and effort behavior. *Journal of Management Accounting Research*, *20* (1), 59–81.

Schreck, P. (2015). Honesty in managerial reporting: How competition affects the benefits and costs of lying. *Critical Perspectives on Accounting*, *27*, 177–188.

Schwartz, M. S. (2004). Effective corporate codes of ethics: Perceptions of code users. *Journal of Business Ethics*, *55* (4), 321–341.

Schwartz, M. S., Dunfee, T. W. & Kline, M. J. (2005). Tone at the top: An ethics code for directors? *Journal of Business Ethics*, *58* (1), 79–100.

Schwering, A. (2016). *Ehrlichkeit in der Budgetierung* (1. Aufl.). Wiesbaden, Springer Gabler.

Schwering, A. (2017). The influence of peer honesty and anonymity on managerial reporting. *Journal of Business Economics*, *87* (9), 1151–1172.

Simons, R. (1995). *Levers of control. How managers use innovative control systems to drive strategic renewal*. Boston, Harvard Business School Press.

Somers, M. J. (2001). Ethical codes of conduct and organizational context: A study of the relationship between codes of conduct, employee behavior and organizational values. *Journal of Business Ethics*, *30* (2), 185–195.

Sprinkle, G. B. & Williamson, M. G. (2006). Experimental research in managerial accounting. *Handbook of Management Accounting Research*, *1*, 415–444.

Stevens, D. E. (2002). The effects of reputation and ethics on budgetary slack. *Journal of Management Accounting Research*, *14* (1), 153–171.

Stevens, D. E. & Thevaranjan, A. (2010). A moral solution to the moral hazard problem. *Accounting, Organizations and Society*, *35* (1), 125–139.

Tayler, W. B. & Bloomfield, R. J. (2011). Norms, conformity, and controls. *Journal of Accounting Research*, *49* (3), 753–790.

Tenbrunsel, A. E., Smith-Crowe, K. & Umphress, E. E. (2003). Building houses on rocks: The role of the ethical infrastructure in organizations. *Social Justice Research*, *16* (3), 285–307.

Towry, K. L. (2003). Control in a teamwork environment—The impact of social ties on the effectiveness of mutual monitoring contracts. *The Accounting Review*, *78* (4), 1069–1095.

Twain, M. (1998). *The Wit and Wisdom of Mark Twain: A Book of Quotations* (1. Aufl.). Mineola, Dover Publications.

Van der Stede, W. A. (2000). The relationship between two consequences of budgetary controls: budgetary slack creation and managerial short-term orientation. *Accounting, Organizations and Society*, *25* (6), 609–622.

Voss, U. & Brettel, M. (2014). The effectiveness of management control in small firms: Perspectives from resource dependence theory. *Journal of Small Business Management*, *52* (3), 569–587.

Waller, W. S. (1988). Slack in participative budgeting: The joint effect of a truth-inducing pay scheme and risk preferences. *Accounting, Organizations and Society*, *13* (1), 87–98.

Warren, D. E., Gaspar, J. P. & Laufer, W. S. (2014). Is formal ethics training merely cosmetic? A study of ethics training and ethical organizational culture. *Business Ethics Quarterly*, *24* (1), 85–117.

Webb, R. A. (2002). The impact of reputation and variance investigations on the creation of budget slack. *Accounting, Organizations and Society*, *27* (4-5), 361–378.

Weiskirchner-Merten, K. (2020). Interdependence, participation, and coordination in the budgeting process. *Business Research*, *13* (1), 247–274.

Weitzman, M. L. (1976). The new Soviet incentive model. *The Bell Journal of Economics*, *7* (1), 251–257.

Wotruba, T. R., Chonko, L. B. & Loe, T. W. (2001). The impact of ethics code familiarity on manager behavior. *Journal of Business Ethics*, *33* (1), 59–69.

Young, S. M. (1985). Participative budgeting: The effects of risk aversion and asymmetric information on budgetary slack. *Journal of Accounting Research*, *23* (2), 829–842.

Zablow, R. J. (2006). Creating and sustaining an ethical workplace. *Journal of Risk Management*, *53* (9), 26.

Zhang, Y. (2008). The effects of perceived fairness and communication on honesty and collusion in a multi-agent setting. *The Accounting Review*, *83* (4), 1125–1146.

Junior Management Science 8(1) (2023) 96-122

Junior Management Science

journal homepage: www.jums.academy

Going-Concern-Opinion – A Critical Analysis of Determinants and Consequences

Going-Concern-Modifizierung des Bestätigungsvermerks – Eine kritische Analyse der Einflussfaktoren und Konsequenzen

Niklas Behne

Universität Hamburg

Abstract

The auditor's assessment and reporting of significant uncertainties regarding the going concern is an important source of information for the stakeholder of an entity. Therefore, this paper analyzes and evaluates the determinants that lead to the so-called going-concern-opinion (GCO) as well as the consequences of a GCO for the company and the auditor. The results indicate that it is more likely for a company to receive a GCO when it is in a weaker economic situation, has a more independent audit committee and has an effective internal control. For companies, the GCO results in a negative market reaction and worse lending conditions. Furthermore, after a GCO an auditor switching is more likely, and the auditor receives less audit fees. In the case of an erroneously omitted GCO the auditor is confronted with the risk of litigation and the loss of his reputation. This systematic overview can be used as basic for regulatory efforts and educate the practice about potential consequences of a GCO.

Zusammenfassung

Die Einschätzung und Berichterstattung des Abschlussprüfers über wesentliche Unsicherheiten in Verbindung mit der Fortführung der Unternehmenstätigkeit stellt für die Stakeholder eines Unternehmens eine wichtige Informationsquelle dar. In der vorliegenden Arbeit werden daher die Determinanten, die zu einer sog. Going-Concern-Opinion (GCO) führen, sowie die möglichen Konsequenzen einer GCO für das Unternehmen und den Abschlussprüfer anhand wissenschaftlicher Forschungsergebnisse analysiert und eingeordnet. Die Ergebnisse zeigen eine größere Wahrscheinlichkeit einer GCO bei Unternehmen in einer finanziell schwächeren Situation, mit einem unabhängigeren Audit Committee bzw. einer wirksamen internen Kontrolle. Für Unternehmen hat die GCO eine negative Marktreaktion sowie schlechtere Konditionen bei der Kreditvergabe zur Folge. Die GCO führt zu einer größeren Wahrscheinlichkeit eines Abschlussprüferwechsels sowie geringeren Prüfungshonoraren beim Abschlussprüfer. Bei einer fälschlicherweise unterlassenen GCO sieht sich der Abschlussprüfer mit der Gefahr von Rechtsstreitigkeiten und einem Reputationsverlust konfrontiert. Dieser systematische Überblick kann als Basis für künftige Regulierungsbestrebungen genutzt werden und die Praxis über die möglichen Folgen einer GCO aufklären.

Keywords: Unternehmensfortführung; Going-Concern-Opinion; Bestätigungsvermerk; Abschlussprüfung.

1. Einleitung

Die Einschätzung über die Fortführung der Unternehmenstätigkeit stellt eine entscheidende Aufgabe der Unternehmensführung in der Rechnungslegung dar und wird durch den Abschlussprüfer[1] auf ihre Angemessenheit überprüft.[2] Die Stakeholder haben ein Interesse daran, durch den Abschlussprüfer über mögliche Unsicherheiten und einen bevorstehenden Zusammenbruch eines Unternehmens informiert

[1] Für eine bessere Lesbarkeit wird in der vorliegenden Arbeit das generische Maskulinum verwendet, das sich zugleich auf die männliche, die weibliche und andere Geschlechteridentitäten bezieht.

[2] Vgl. Frey (2014, S. 1); Heinrichs (2018, S. 1).

DOI: https://doi.org/10.5282/jums/v8i1pp96-122

zu werden, auch wenn der Abschlussprüfer nicht für die Vorhersage zukünftiger Ereignisse zuständig ist.[3] Neben der Finanzkrise in den Jahren 2008 und 2009 gewann die Annahme der Unternehmensfortführung durch die COVID-19-Pandemie im Jahr 2020 an Bedeutung, da sich Unternehmen mit Einschränkungen und wirtschaftlichen Folgen konfrontiert sahen.[4] Bei der Annahme der Unternehmensfortführung handelt es sich um eine zukunftsorientierte Aussage, die eine Prognose über die künftige Entwicklung eines Unternehmens abgibt, weshalb die Beurteilung mit Unsicherheiten verbunden ist und eine der schwierigsten Problemfelder für den Abschlussprüfer darstellt.[5]

Wird die Annahme der Unternehmensfortführung durch den Abschlussprüfer als nicht angemessen angesehen oder gibt es wesentliche Zweifel an der Fortführung der Unternehmenstätigkeit, wird eine Going-Concern-Modifizierung des Bestätigungsvermerks vorgenommen.[6] Gegenstand dieser Arbeit ist die kritische Analyse der Determinanten, die einen Einfluss auf die Erteilung einer Going-Concern-Modifizierung haben. Ebenso werden die Konsequenzen und der informative Wert für die Stakeholder einer solchen Modifizierung genauer beleuchtet.

Im nachfolgenden Kapitel wird ein Überblick über die regulatorischen und theoretischen Grundlagen gegeben, bevor in Kapitel 3 die Einflussfaktoren der Going-Concern-Modifizierung untersucht werden. Hierbei wird zwischen den Faktoren beim zu prüfenden Unternehmen, den Faktoren beim Abschlussprüfer und der Beziehung zwischen Abschlussprüfer und Unternehmen differenziert. Anschließend erfolgt in Kapitel 4 die Analyse der Konsequenzen der Going-Concern-Modifizierung sowohl hinsichtlich der Wirkung beim zu prüfenden Unternehmen als auch dem Effekt beim Abschlussprüfer. Die aufgeführten Ergebnisse werden in Kapitel 5 einer kritischen Würdigung unterzogen und abschließend im Fazit zusammengefasst.

2. Regulatorische und theoretische Grundlagen

2.1. Going-Concern-Prämisse und Verantwortlichkeiten

Einer der wesentlichen Grundsätze der Rechnungslegung ist das Prinzip der Annahme der Unternehmensfortführung, auch Going-Concern-Prämisse[7] genannt, das sowohl in nationalen als auch in internationalen Rechnungslegungsnormen Berücksichtigung findet.[8] Nach den handelsrechtlichen Vorgaben des *§ 252 Abs. 1 Nr. 2 HGB* sind die im Jahresabschluss ausgewiesenen Vermögensgegenstände und Schulden so zu bewerten, dass von einer Fortführung der Unternehmenstätigkeit auszugehen ist, solange keine tatsächlichen oder rechtlichen Gegebenheiten der Going-Concern-Prämisse widersprechen.[9] Wird die Going-Concern-Prämisse nicht erfüllt, erfolgt der Ansatz der Bilanzposten nicht wie oben beschrieben mit den Fortführungswerten, sondern mit den Zerschlagungswerten.[10] Die International Accounting Standards (IAS) erfordern ebenfalls die Aufstellung der Jahresabschlüsse auf Grundlage der Going-Concern-Prämisse, außer das Management plant die Auflösung des Unternehmens bzw. die Einstellung der Geschäftstätigkeit oder hat keine weitere realistische Alternative als entsprechend zu handeln.[11] Sollten der Unternehmensführung Umstände bekannt sein, die zu erheblichen Zweifeln an der Fortführung der Unternehmenstätigkeit führen, muss im Jahresabschluss auf diese Unsicherheiten eingegangen werden.

Die Ermessensentscheidung über die Fähigkeit der Unternehmensfortführung obliegt den gesetzlichen Vertretern.[12] Sie entscheiden, ob tatsächliche oder rechtliche Gegebenheiten der Fortführung der Geschäftstätigkeit entgegenstehen.[13] Zu den tatsächlichen Gegebenheiten gehören hauptsächlich wirtschaftliche Schwierigkeiten, wie bspw. die Unmöglichkeit der Beschaffung neuer Finanzmittel für Investitionen, Engpässe bei der Beschaffung von Vorräten sowie der Wegfall von wesentlichen Kreditgebern, Zulieferern oder Kunden. Unter den rechtlichen Gegebenheiten sind z. B. die Eröffnung des Insolvenzverfahrens aufgrund von Zahlungsunfähigkeit, drohender Zahlungsunfähigkeit bzw. Überschuldung sowie die gesetzlichen Vorschriften, die im Weiteren zur Einstellung der Geschäftstätigkeit führen, zu verstehen. Rufen Ereignisse oder Gegebenheiten im Einzelnen oder insgesamt bedeutsame Zweifel an der Unternehmensfortführung hervor, müssen die gesetzlichen Vertreter im Abschluss darüber berichten und eindeutig darauf hinweisen, dass es wesentliche Unsicherheiten im Zusammenhang mit der Fortführung der Unternehmenstätigkeit gibt, damit der Abschluss ein den tatsächlichen Verhältnissen entsprechendes Bild der Vermögens-, Finanz- und Ertragslage (VFE-Lage) darstellt.[14] Sollten die Informationen im Lagebericht offengelegt werden, ist im Abschluss ein eindeutiger Verweis auf die Angaben im Lagebericht aufzunehmen.

Die Beurteilung der Going-Concern-Prämisse ist insbesondere während der COVID-19-Pandemie für Unternehmen in den Vordergrund gerückt, da das Management die Unternehmensfortführung nicht nur unter Beachtung der rechtlichen Einschränkungen, sondern auch unter dem Aspekt der wirtschaftlichen Folgen einschätzen muss.[15] Die Bewertung der Lage ist dabei aufgrund zahlreicher Ungewissheiten, wie

[3] Vgl. Geiger, Gold und Wallage (2021, S. 3).
[4] Vgl. Pföhler (2021, S. 354); Ruhnke und Frey (2015, S. 1).
[5] Vgl. Ruhnke und Frey (2015, S. 1).
[6] Vgl. Heinrichs (2018, S. 1).
[7] In der Literatur gibt es für den Begriff der Going-Concern-Prämisse eine Vielzahl an Termini, wie bspw. Prinzip bzw. Grundsatz der Unternehmensfortführung. In dieser Arbeit wird einheitlich der Begriff Going-Concern-Prämisse verwendet.
[8] Vgl. Adam (2007, S. 24 u. 63); Eickes (2014, S. 4); Marten, Quick und Ruhnke (2020, S. 597).
[9] Vgl. §252 Abs. 1 Nr. 2 HGB (2020).
[10] Vgl. Sarburg und Mengwasser (2017, S. 1179).
[11] Vgl. hierfür und nachfolgend IAS 1, Tz. 25.
[12] Vgl. IDW PS 270 n. F. (2021, Tz. 5); Pföhler (2021, S. 355).
[13] Vgl. hierfür und nachfolgend Eickes (2014, S. 119-120); Marten et al. (2020, S. 597).
[14] Vgl. hierfür und nachfolgend IDW PS 270 n. F. (2021, Tz. 9 u. Tz. 23); Marten et al. (2020, S. 600).
[15] Vgl. hierfür und nachfolgend Pföhler (2021, S. 354).

bspw. der Länge eines Lockdowns, mit Unsicherheiten verbunden. Das Institut der Wirtschaftsprüfer (IDW) geht in seinem fachlichen Hinweis von einem wertbegründenden Ereignis aus, da die Ausweitung des Virus erst im Januar 2020 erfolgte, sodass bilanzielle Konsequenzen erst in Abschlüsse mit einem Bilanzstichtag nach dem 31.12.2019 aufzunehmen sind.[16] Vom Gesetzgeber wurden zur Bewältigung der Pandemie diverse Maßnahmen ergriffen, wozu u. a. das COVID-19-Insolvenzaussetzunggesetz (COVInsAG) zählt.[17] Durch dieses Gesetz wurde der Prognosezeitraum für die Überschuldungsprüfung auf vier Monate verkürzt und für Unternehmen, bei denen die Insolvenzreife auf den Auswirkungen der COVID-19-Pandemie beruht, wurde die Insolvenzantragspflicht ausgesetzt.

Während die Unternehmensleitung für die Einschätzung der Going-Concern-Prämisse verantwortlich ist, hat der gesetzliche Abschlussprüfer die Angemessenheit der Going-Concern-Prämisse zu beurteilen und eine Schlussfolgerung zu ziehen, ob eine wesentliche Unsicherheit hinsichtlich der Fortführung der Unternehmenstätigkeit besteht.[18] Hierbei ist zu beachten, dass sich die Zusicherung des Fortbestands des Unternehmens bzw. eine mögliche Insolvenzantragspflicht nicht auf den Zuständigkeitsbereich des Abschlussprüfers erstreckt.[19] Identifiziert der Abschlussprüfer während der Prüfungsplanung oder -durchführung relevante finanzwirtschaftliche, betriebliche oder sonstige Ereignisse bzw. Gegebenheiten, die eine Unternehmensfortführung in Frage stellen, dann hat der Abschlussprüfer weitere Prüfungshandlungen durchzuführen sowie ausreichend geeignete Prüfungsnachweise einzuholen, um die Angemessenheit der Going-Concern-Prämisse und der vom Unternehmen getätigten Angaben im Abschluss abschließend beurteilen zu können.[20] Dabei ist es möglich, dass auf der einen Seite Ereignisse bzw. Gegebenheiten alleine oder kombiniert zu wesentlichen Unsicherheiten führen, aber auf der anderen Seite Faktoren die relevanten Ereignisse bzw. Gegebenheiten abmildern.[21] Der Abschlussprüfer hat daher die Bedeutung und Wirkungsweise einzelner Faktoren in seiner Einschätzung zu berücksichtigen.

Zur Beurteilung der Unternehmensfortführung wird grundsätzlich auf den Bilanzstichtag abgestellt, wobei Ereignisse, die nach dem Abschlussstichtag auftreten und die Going-Concern-Prämisse beeinflussen, bis zum Zeitpunkt der Erteilung des Bestätigungsvermerks berücksichtigt werden.[22] Die gesetzlichen Vertreter müssen nach *IAS 1.26* alle verfügbaren Informationen in einem Zeitraum von mindestens zwölf Monaten ab dem Abschlussstichtag in ihrer Entscheidung über die Going-Concern-Prämisse beachten.[23] Der Abschlussprüfer hat bei der Einschätzung über die Angemessenheit der Annahme der Unternehmensfortführung nach *ISA 570 (Revised)* sowie *IDW PS 270 n. F.* den Zeitraum zu berücksichtigen, der den gesetzlichen Vertretern durch die Rechnungslegungsgrundsätze oder andere Vorschriften vorgegeben wird.[24] Sollte der Zeithorizont weniger als zwölf Monate betragen, hat der Abschlussprüfer die Unternehmensleitung zu einer Ausweitung des Zeitraumes der Einschätzung aufzufordern. Treten bei einer Befragung der gesetzlichen Vertreter Ereignisse oder Gegebenheiten auf, die nach dem Zeitraum der Going-Concern-Prämisse liegen und zu bedeutsamen Zweifeln an der Fortführung der Unternehmenstätigkeit führen, kann der Abschlussprüfer die Unternehmensführung zu einer Beurteilung der Bedeutung dieser Ereignisse oder Gegebenheiten auffordern und u. U. erweiterte Prüfungshandlungen durchführen.[25] Bei der Beurteilung der Going-Concern-Prämisse gilt es zudem zu beachten, dass es sich um eine Prognose handelt, die eine gewisse Unsicherheit aufweist und der Unsicherheitsgrad zunimmt, desto weiter sich das Ereignis oder die Gegebenheit in der Zukunft befindet.[26] Darüber hinaus handelt es sich bei der Einschätzung, ob ein Ereignis oder eine Gegebenheit der Going-Concern-Prämisse entgegensteht, um eine subjektive Entscheidung, die einen gewissen Ermessensspielraum mit sich bringt.[27]

2.2. Berichterstattung des Abschlussprüfers

Der Abschlussprüfer verfasst nach Beendigung der Prüfung einen Bestätigungsvermerk und stellt in diesem das Ergebnis sowie Gegenstand, Art und Umfang seiner Prüfung dar.[28] Werden während der Prüfung Risiken identifiziert, die den Fortbestand des Unternehmens in Frage stellen, so muss der Abschlussprüfer im Bestätigungsvermerk gesondert auf diese Thematik eingehen.[29] Die Berichterstattung über die Einschätzung der Going-Concern-Prämisse durch den Abschlussprüfer kann dabei unterschiedliche Ausprägungen annehmen.[30] Ein Bestätigungsvermerk ohne Ergänzung (Unqualified Opinion) ist grundsätzlich dann auszustellen, wenn die von der Unternehmensführung getätigte Going-Concern-Prämisse durch den Abschlussprüfer als angemessen beurteilt wird und keine wesentlichen gegenteiligen Informationen vorliegen.

[16] Vgl. IDW (Hrsg.) (2020, S. 2).

[17] Vgl. hierfür und nachfolgend §1 Abs. 1 u. §4 Abs. 1 Satz 1 COVInsAG, COVInsAG (2021).

[18] Vgl. IDW PS 270 n. F. (2021, Tz. 10); ISA 570 (Revised) (2015, Tz. 6); Meyer und Davids (2018, S. 265).

[19] Vgl. §317 Abs. 4a HGB (2020); IDW PS 270 n. F. (2021, Tz. 12 u. A12).

[20] Vgl. ISA 570 (Revised) (2015, Tz. 9 i. V. m. Tz. 16-17); Marten et al. (2020, S. 602-604).

[21] Vgl. hierfür und nachfolgend Marten et al. (2020, S. 602).

[22] Vgl. Marten et al. (2020, S. 598-599).

[23] Vgl. IAS 1, Tz. 26.

[24] Vgl. hierfür und nachfolgend IDW PS 270 n. F. (2021, Tz. 18); ISA 570 (Revised) (2015, Tz. 13).

[25] Vgl. IDW PS 270 n. F. (2021, Tz. 20 i. V. m. Tz. A16); ISA 570 (Revised) (2015, Tz. 15).

[26] Vgl. Adam (2007, S. 28); Eickes (2014, S. 117); IDW PS 270 n. F. (2021, Tz. A16).

[27] Vgl. Eickes (2014, S. 117).

[28] Vgl. §322 Abs. 1 HGB (2020).

[29] Vgl. §322 Abs. 2 Satz 3 HGB (2020).

[30] Vgl. hierfür und nachfolgend Heinrichs (2018, S. 20-23); IDW PS 270 n. F. (2021, Tz. 22 f.); ISA 570 (Revised) (2015, Tz. 17 f.); Marten et al. (2020, S. 608-609).

Zu einer Modifizierung des Bestätigungsvermerks kommt es dagegen, wenn eine Einwendung oder ein Prüfungshemmnis vorhanden ist.[31] Die Modifizierung kann neben der Einschränkung oder Versagung des Prüfungsurteils auch in Form einer Erklärung der Nichtabgabe eines Prüfungsurteils erfolgen.[32] Hierbei ist entscheidend, ob die Einwendungen bzw. Prüfungshemmnisse wesentlich, aber nicht umfassend (Einschränkung des Prüfungsurteils) oder wesentlich und umfassend (Versagung des Prüfungsvermerks) sind.[33] Ein Versagungsvermerk (Adverse Opinion) wird u. a. dann erteilt, wenn der Abschlussprüfer die Going-Concern-Prämisse des Managements als nicht angemessen erachtet.[34] Könnten die Prüfungshemmnisse nach Einschätzung des Abschlussprüfers wesentlich und umfassend sein und es war ihm nicht möglich, sich ein Prüfungsurteil zu bilden, ist die Erklärung der Nichtabgabe eines Prüfungsurteils (Disclaimer of Opinion) geboten.[35]

Liegen bei einer angemessenen Going-Concern-Prämisse relevante Ereignisse oder Gegebenheiten vor, aber eine wesentliche Unsicherheit ist nicht gegeben, weil bspw. die vom Management eingeleiteten Maßnahmen als geeignet eingeschätzt werden, dann hat der Abschlussprüfer bei einer angebrachten Darstellung der Ereignisse oder Gegebenheiten, unter Bezugnahme auf die unternommenen Gegenmaßnahmen, einen uneingeschränkten Bestätigungsvermerk ohne Zusatz abzugeben.[36] Dahingegen erfolgt bei einem Abschluss mit nicht angemessenen Angaben die Modifizierung des Prüfungsurteils in Form eines eingeschränkten Bestätigungsvermerks (Qualified Opinion) oder Versagungsvermerks. Kommt der Abschlussprüfer zu dem Schluss, dass die Going-Concern-Prämisse der Unternehmensführung als angemessen beurteilt werden kann und es bestehen wesentliche Unsicherheiten, dann ist entscheidend, ob das Unternehmen auf diese wesentlichen Unsicherheiten in seinen Abschlussangaben entsprechend eingegangen ist.[37] Wenn die Angaben zu der wesentlichen Unsicherheit im Jahresabschluss adäquat aufgezeigt wurden, wird der Bestätigungsvermerk nicht modifiziert, jedoch um einen Absatz mit der Überschrift „Wesentliche Unsicherheit im Zusammenhang mit der Fortführung der Unternehmenstätigkeit“[38] ergänzt, der darauf hinweist, dass eine wesentliche Unsicherheit in Bezug auf die Unternehmensfortführung existiert. Im Gegensatz dazu wird bei einer unzureichenden Darstellung der wesentlichen Unsicherheiten eine Modifizierung des Bestätigungsvermerks durch eine Einschränkung des Bestätigungsvermerks oder einen Versagungsvermerk vorgenommen.[39] Zeigt sich bei den gesetzlichen Vertretern nach Aufforderung durch den Abschlussprüfer eine fehlende Bereitschaft, ihre Einschätzung vorzunehmen oder den der Einschätzung zugrunde liegenden Zeitraum zu erweitern, kann das Prüfungsurteil eingeschränkt werden oder der Abschlussprüfer erklärt die Nichtabgabe eines Prüfungsurteils.[40]

Zusätzlich zu den Auswirkungen im Bestätigungsvermerk hat der Abschlussprüfer die festgestellten Ereignisse oder Gegebenheiten den mit der Überwachung betrauten Verantwortlichen zu kommunizieren.[41] Ferner werden nach den deutschen Normen in *§ 321 Abs. 1 HGB* vom Abschlussprüfer Angaben im Prüfungsbericht über wesentliche Unsicherheiten und Entwicklungsbeeinträchtigungen gefordert.[42]

2.3. Definition der Unabhängigkeit und finanzieller Faktoren

In der Prüfungsforschung gibt es verschiedene Ansätze, die Prüfungsqualität zu definieren. Nach DeAngelo (1981b) wird die Prüfungsqualität als die gemeinschaftliche Wahrscheinlichkeit, dass der Abschlussprüfer sowohl einen Fehler im Rechnungslegungssystem eines Unternehmens entdeckt als auch über diesen berichtet, definiert.[43] Die Aufdeckung von Verstößen wird dabei u. a. durch die technischen Möglichkeiten des Abschlussprüfers und die angewandten Prüfungsverfahren beeinflusst. In der Definition wird die bedingte Wahrscheinlichkeit, dass der Abschlussprüfer über einen Verstoß berichtet, als ein Maß für die Unabhängigkeit des Abschlussprüfers angesehen. Die Definition verbindet somit die Fähigkeit des Abschlussprüfers, eine wesentliche Unsicherheit zu entdecken, mit der Bereitschaft, über diese zu berichten.[44] Leffson (1988) berücksichtigt in seiner Definition für ein vertrauenswürdiges Prüfungsurteil die Urteilsfähigkeit und die Urteilsfreiheit des Urteilenden sowie die sachgerechte Urteilsbildung.[45] Dabei wird die Urteilsfähigkeit durch die fachliche Qualifikation des Abschlussprüfers dargestellt, während die Urteilsfreiheit dann vorliegt, wenn das Prüfungsurteil vom Abschlussprüfer ohne jegliche Einwirkungen und somit sowohl unabhängig als auch unbefangen erteilt wird.

Die Urteilsfreiheit und damit die Unabhängigkeit kann durch verschiedene Situationen gefährdet sein.[46] Die Gefährdung kann bspw. gegeben sein, wenn personelle Verflechtungen, finanzielle Interessen, persönliche Beziehungen oder eine Verbindung der Prüfungs- und Beratungstätigkeit vorliegen. Die Einschränkung der Unabhängigkeit des

[31] Vgl. IDW PS 405 (2017, Tz. 9).
[32] Vgl. IDW PS 405 (2017, Tz. 4).
[33] Vgl. IDW PS 405 (2017, Tz. 10 u. 12).
[34] Vgl. IDW PS 270 n. F. (2021, Tz. 28); ISA 570 (Revised) (2015, Tz. 21); Marten et al. (2020, S. 608).
[35] Vgl. IDW PS 405 (2017, Tz. 13-14); Marten et al. (2020, S. 611).
[36] Vgl. hierfür und nachfolgend IAS 1 (2020, Tz. 122); IDW PS 270 n. F. (2021, Tz. 26); ISA 570 (Revised) (2015, Tz. 20); Marten et al. (2020, S. 609-612).
[37] Vgl. hierfür und nachfolgend IDW PS 270 n. F. (2021, Tz. 24 u. 29); ISA 570 (Revised) (2015, Tz. 19 u. 22); Marten et al. (2020, S. 610).
[38] IDW PS 270 n. F. (2021, Tz. 29).
[39] Vgl. hierfür und nachfolgend IDW PS 270 n. F. (2021, Tz. 31 f.); ISA 570 (Revised) (2015, Tz. 23); Marten et al. (2020, S. 612).
[40] Vgl. IDW PS 270 n. F. (2021, Tz. A43); Marten et al. (2020, S. 611).
[41] Vgl. ISA 570 (Revised) (2015, Tz. 25); IDW PS 270 n. F. (2021, Tz. 34).
[42] Vgl. IDW PS 270 n. F. (2021, Tz. 35); §321 Abs. 1 HGB (2020); Marten et al. (2020, S. 612).
[43] Vgl. hierfür und nachfolgend DeAngelo (1981b, S. 186).
[44] Vgl. Marten et al. (2020, S. 613).
[45] Vgl. hierfür und nachfolgend Leffson (1988, S. 61); Marten et al. (2020, S. 241).
[46] Vgl. hierfür und nachfolgend Brösel, Freichel, Toll und Buchner (2015, S. 82-83).

Abschlussprüfers aufgrund finanzieller Einflüsse kann sich unterschiedlich bemerkbar machen. Generell besteht zwischen dem Mandanten und dem Abschlussprüfer ein sog. bilaterales Monopol.[47] Beide Parteien haben Anreize, ihre Geschäftsbeziehung weiterhin zu verfolgen, denn der Abschlussprüfer kann in Folgeprüfungen ein Honorar oberhalb der Prüfungskosten verlangen und somit Quasi-Renten erzielen und der Mandant vermeidet Transaktionskosten durch einen Wechsel des Abschlussprüfers. Der Mandant verfolgt zudem das Ziel der Veröffentlichung von Informationen in seinem Sinne, da die Beurteilung des Abschlussprüfers den Unternehmenswert beeinflussen kann und der Abschlussprüfer die Einnahmen, die er für die Durchführung der Prüfung vom Mandanten erhält, aufrechterhalten möchte.[48] Wenn der Abschlussprüfer dem Druck des Mandanten für die Erteilung eines uneingeschränkten Bestätigungsvermerks nicht standhalten kann, wird die Unabhängigkeit des Abschlussprüfers beeinträchtigt und die Prüfungsqualität reduziert.[49] Die Ausstellung einer Going-Concern-Opinion (GCO)[50] hat daher objektiv und unabhängig von der Absicht des Mandanten, ein uneingeschränktes Prüfungsurteil zu erhalten, zu erfolgen.[51] Unter dieser Annahme korreliert die Bereitschaft des Abschlussprüfers, eine GCO auszustellen, positiv mit der Unabhängigkeit des Abschlussprüfers.

Da die Unabhängigkeit des Abschlussprüfers und somit auch die Prüfungsqualität nicht direkt beobachtet werden kann, werden in der Prüfungsforschung Ersatzgrößen, sog. Proxys, wie bspw. Bestätigungsvermerke mit einer GCO, zur Operationalisierung und Messung der Prüfungsqualität eingesetzt.[52] GCOs können insbesondere bei der Prüfung von wahrgenommenen Bedrohungen für die Prüfungsqualität, wie z. B. Nichtprüfungshonorare, die Größe des Mandanten und die Beschäftigungsdauer, verwendet werden.[53] Die GCO stellt dabei ein direktes Maß für die Prüfungsqualität dar, weil der Abschlussprüfer für den Bestätigungsvermerk verantwortlich zeichnet und dieser unter seinem Einfluss und seiner Kontrolle entsteht. Zudem ist die GCO gleichzeitig eine diskrete Messgröße, über deren Messung eine relativ hohe Einigkeit besteht und die einen verhältnismäßig geringen Messfehler aufweist.

Viele Studien verwenden zur Untersuchung des Effekts verschiedener Determinanten auf die GCO Unternehmen, die sich in einer finanziellen Schieflage befinden (Financially Distressed Companies) und setzen diese als Kontrollgruppe ein.[54] Diese Vorgehensweise ist auf vorangegangene Untersuchungen zurückzuführen, in denen festgestellt wurde, dass bei empirischer Betrachtung praktisch keine GCOs an Unternehmen ausgestellt wurden, die sich nicht in finanziellen Schwierigkeiten befanden.[55] Durch die Differenzierung der finanziellen Lage kann die statistische Schätzung des Modells durch die Erhöhung der gruppeninternen Homogenität der Stichprobe verbessert und somit eine Steigerung der Aussagekraft erreicht werden.[56] Darüber hinaus werden damit Mandate erfasst, die vom Abschlussprüfer ein hohes Maß an Urteilsvermögen erfordern.[57] Die Identifikation der finanziell gestressten Unternehmen erfolgt anhand verschiedener Parameter, wobei sich die verwendeten Kennzahlen in den Studien unterscheiden.[58] Ein Unternehmen wird bspw. als finanziell belastet eingestuft, wenn ein negatives Eigenkapital, ein negativer operativer Cashflow, ein negatives Working Capital, ein negatives EBIT oder ein Jahresfehlbetrag im zurückliegenden Jahr bzw. den zwei vorangegangenen Jahren erkannt wird. Einige Studien definieren den Rahmen für die Einordnung strikter, sodass zwei Parameter erfüllt sein müssen und sich das Unternehmen somit in einer größeren finanziellen Schieflage befindet.[59]

Im Nachfolgenden werden diverse finanzielle Faktoren als Determinanten für eine GCO aufgezeigt, die in den einzelnen Studien unterschiedlich eingesetzt werden. Unter der Profitabilität, im Englischen auch bekannt als Return-on-Assets (ROA), wird der Quotient aus dem Nettoertrag und dem Gesamtvermögen, bestehend aus dem Eigenkapital und dem Fremdkapital, verstanden.[60] Die Liquidität eines Unternehmens wird bspw. anhand des Verhältnisses des Umlaufvermögens zu den kurzfristigen Verbindlichkeiten gemessen und zeigt, ob ein Unternehmen seine kurzfristigen Verbindlichkeiten bedienen kann.[61] Das Leverage beschreibt die Finanzierungsstruktur eines Unternehmens und wird z. B. durch das Verhältnis der Gesamtverbindlichkeiten, seltener nur die Summe der langfristigen Verbindlichkeiten, zur Summe der Vermögenswerte ermittelt.[62] Es kann als ein Indikator zur Bestimmung der Fähigkeit eines Unternehmens, seine kurzfristigen und langfristigen Schulden zu bedienen, eingesetzt werden. Dabei zeigt ein höheres Level an Leverage an, dass die Mittelherkunft hauptsächlich aus Darlehen resultiert.

[47] Vgl. hierfür und nachfolgend Marten et al. (2020, S. 242).
[48] Vgl. DeAngelo (1981a, S. 117).
[49] Vgl. DeFond und Zhang (2014, S. 284 u. 287).
[50] Die Going-Concern-Modifizierung ist im Englischen auch als Going-Concern-Opinion bekannt, daher wird im Folgenden die Abkürzung GCO verwendet.
[51] Vgl. hierfür und nachfolgend DeFond, Raghunandan und Subramanyam (2002, S. 1248-1249).
[52] Vgl. J. R. Francis (2011, S. 129); Marten et al. (2020, S. 613).
[53] Vgl. hierfür und nachfolgend DeFond und Zhang (2014, S. 287).
[54] Vgl. Geiger et al. (2021, S. 24).
[55] Vgl. Carcello und Neal (2000, S. 457); McKeown, Mutchler und Hopwood (1991, S. 10).
[56] Vgl. Hopwood, McKeown und Mutchler (1994, S. 412).
[57] Vgl. Ratzinger-Sakel (2013, S. 137).
[58] Vgl. hierfür und nachfolgend Hopwood et al. (1994, S. 417); McKeown et al. (1991, S. 6); Ratzinger-Sakel (2013, S. 138).
[59] Vgl. Berglund, Eshleman und Guo (2018, S. 9); Carey, Geiger und O'connell (2008, S. 67); Ratzinger-Sakel (2013, S. 138).
[60] Vgl. Kida (1980, S. 510); Koh und Killough (1990, S. 186); Reichelt und Wang (2010, S. 660).
[61] Vgl. Kida (1980, S. 510); Parker, Peters und Turetsky (2005, S. 14); Simamora und Hendarjatno (2019, S. 146).
[62] Vgl. hierfür und nachfolgend Averio (2021, S. 157); G. V. Krishnan und Wang (2015, S. 144); Parker et al. (2005, S. 13); Reichelt und Wang (2010, S. 664).

3. Determinanten der Going-Concern-Modifizierung

3.1. Faktoren beim zu prüfenden Unternehmen

3.1.1. Finanzielle Kennzahlen des Jahresabschlusses und Marktes

In der Prüfungsforschung befassen sich zahlreiche Studien mit dem Zusammenhang zwischen finanziellen Kennzahlen des Jahresabschlusses und der Erteilung einer GCO. Bei der Betrachtung der Profitabilität eines Unternehmens stellen u. a. Kida (1980), Koh und Killough (1990) sowie Averio (2021) einen negativen Einfluss der Profitabilität auf die Wahrscheinlichkeit der Ausstellung einer GCO fest.[63] Dies bedeutet, dass es für ein Unternehmen wahrscheinlicher ist, einen modifizierten Bestätigungsvermerk zu erhalten, wenn das Unternehmen weniger profitabel ist.

Die Untersuchung der Verbindung zwischen der Liquidität eines Unternehmens und einer GCO ist ebenfalls Gegenstand verschiedener Studien.[64] Einen negativen Effekt der Liquidität auf die GCO ermitteln sowohl Koh und Killough (1990) als auch Raghunandan und Rama (1995) und im Weiteren Averio (2021). Die Wahrscheinlichkeit eine GCO zu erhalten ist somit größer, wenn die Liquidität des Unternehmens geringer ist. Dahingegen kommen Simamora und Hendarjatno (2019) in ihrer Analyse von Unternehmen der Indonesia Stock Exchange zu dem Ergebnis, dass das Niveau der Liquidität keinen Einfluss auf die GCO hat.[65]

Ein signifikant positiver Zusammenhang zwischen dem Leverage eines Unternehmens und einer GCO wird bspw. von Raghunandan und Rama (1995), Ye, Carson und Simnett (2011) und G. V. Krishnan und Wang (2015) konstatiert.[66] Weist ein Unternehmen ein höheres Level an Leverage auf, ist dies nicht nur mit einem höheren Investitionsrisiko, sondern auch einer größeren Wahrscheinlichkeit eine GCO zu erhalten, verbunden. Durch ein hohes Leverage werden Zweifel an der Fortführung der Geschäftstätigkeit eines Unternehmens erhoben, da die finanziellen Mittel zur Finanzierung der Schulden eingesetzt werden und dies mit einer gleichzeitigen Reduzierung der Mittel für die laufende Betriebstätigkeit einhergeht.[67]

Die Analysen von McKeown et al. (1991), Geiger und Raghunandan (2001), Knechel und Vanstraelen (2007) und G. V. Krishnan und Wang (2015) identifizieren darüber hinaus einen signifikant negativen Effekt zwischen der Größe eines Unternehmens und dem Empfang einer GCO.[68] Für ein Unternehmen ist es damit wahrscheinlicher, eine GCO zu erhalten, wenn das Unternehmen kleiner ist. Mutchler, Hopwood und McKeown (1997) begründet das damit, dass bei größeren Unternehmen der Abschlussprüfer von einer Erholung des Unternehmens überzeugt ist oder aber Anlass zur Sorge vor einem Zusammenbruch des Unternehmens, herbeigeführt durch die Erteilung einer GCO, hat.[69] Zu einer anderen Erkenntnis kommen Carcello, Vanstraelen und Willenborg (2009), die bei ihrer Studie in Belgien keine signifikante Verbindung zwischen der Unternehmensgröße und dem Vorliegen einer GCO nachweisen können.[70] Ebenso stellt Achyarsyah (2016) für den indonesischen Markt keinen signifikanten Einfluss der Unternehmensgröße auf die GCO fest.[71] Der Autor argumentiert zudem, dass die Größe eines Unternehmens nicht als ein beeinflussender Faktor in Frage komme, da diese keinen Anhaltspunkt bei der Erstellung einer GCO darstelle.

Neben den finanziellen Kennzahlen des Jahresabschlusses gibt es auch weitere finanzielle Maßzahlen, die mit der Ausstellung einer GCO verbunden sind. Hierzu zählen bspw. Marktvariablen, wie die Renditevolatilität und die (branchenbereinigten) Renditen. Die Analyse von DeFond et al. (2002) weist auf eine höhere Wahrscheinlichkeit für die Erteilung einer GCO bei einer höheren Renditevolatilität hin.[72] Diese Einschätzung wird durch die Ergebnisse von sowohl Goh, Krishnan und Li (2013) als auch G. V. Krishnan und Wang (2015) unterstützt, die bei einer größeren Volatilität die Wahrscheinlichkeit, eine GCO zu erhalten, höher bewerten.[73] Ebenfalls zeigt die Studie von Griffin und Lont (2010), dass GCOs positiv mit Unternehmen verbunden sind, die aufgrund einer höheren Volatilität ein größeres Risiko aufweisen, sowie negativ mit Unternehmen mit geringerer Rendite.[74] Zu beachten gilt bei diesen Kennzahlen, dass in der Literatur nicht explizit darauf eingegangen wird, ob diese Kennzahlen vom Abschlussprüfer bei seiner Entscheidung über die Modifizierung des Bestätigungsvermerks berücksichtigt werden.[75] Fraglich ist daher, ob der Abschlussprüfer diese Informationen tatsächlich in seine Urteilsbildung einbezieht oder ob sich die Zusammenhänge durch eine Korrelation mit dem Bestätigungsvermerk ergeben.[76]

3.1.2. Unternehmensstrategie und Finanzberichterstattung

Neben den gegenteiligen Informationen (Contrary Information), die die Fortführung der Unternehmenstätigkeit in Frage stellen, gibt es auch die sog. mildernden Faktoren (Mitigating Factors), die die Wirkung der gegenteiligen Informationen abschwächen.[77] Mutchler et al. (1997) analysieren die gegenteiligen Informationen und mildernden Faktoren bei Unternehmen, die in der Zeit von 1974 bis 1985 insolvent gegangen sind.[78] Die Autoren ermitteln einen signifi-

[63] Vgl. hierfür und nachfolgend Averio (2021, S. 161); Kida (1980, S. 513); Koh und Killough (1990, S. 186).

[64] Vgl. hierfür und nachfolgend Averio (2021, S. 161); Koh und Killough (1990, S. 186); Raghunandan und Rama (1995, S. 57).

[65] Vgl. Simamora und Hendarjatno (2019, S. 154).

[66] Vgl. hierfür und nachfolgend G. V. Krishnan und Wang (2015, S. 154); Raghunandan und Rama (1995, S. 58); Ye et al. (2011, S. 140-141).

[67] Vgl. Achyarsyah (2016, S. 6772).

[68] Vgl. hierfür und nachfolgend Geiger und Raghunandan (2001, S. 191); Knechel und Vanstraelen (2007, S. 126); G. V. Krishnan und Wang (2015, S. 154); McKeown et al. (1991, S. 9).

[69] Vgl. Mutchler et al. (1997, S. 306-307).

[70] Vgl. Carcello et al. (2009, S. 1398).

[71] Vgl. hierfür und nachfolgend Achyarsyah (2016, S. 6779).

[72] Vgl. DeFond et al. (2002, S. 1261).

[73] Vgl. Goh et al. (2013, S. 982); G. V. Krishnan und Wang (2015, S. 154).

[74] Vgl. Griffin und Lont (2010, S. 26).

[75] Vgl. Carson et al. (2013, S. 358).

[76] Vgl. Geiger et al. (2021, S. 27).

[77] Vgl. Mutchler et al. (1997, S. 295).

[78] Vgl. hierfür und nachfolgend Mutchler et al. (1997, S. 302 u. 306).

kanten Effekt von gegenteiligen Informationen in Form von stark negativen Neuigkeiten und schließen daraus, dass der Abschlussprüfer bei seiner Entscheidung gegebenenfalls andere Faktoren berücksichtigt, als solche, die in Verbindung zu einer Insolvenz stehen. Die Ergebnisse für die mildernde Faktoren stellen sich als insignifikant heraus. Begründet wird dies von Mutchler et al. (1997) mit einem möglichen Prüferkonservatismus.

Strategische Entscheidungen vom Management, die der Öffentlichkeit bekannt sind, können ebenfalls Auswirkungen auf die Wahrscheinlichkeit der Erteilung einer GCO haben, wie Behn, Kaplan und Krumwiede (2001) in ihrer Studie von Unternehmen mit einer GCO im Zeitraum 1992 bis 1995 zeigen.[79] Wenn das Management Pläne hat, neue Aktien auszugeben, steht das in einem signifikant negativen Zusammenhang mit der Wahrscheinlichkeit, eine GCO zu erhalten. Dies gilt ebenso für Planungen im Hinblick auf die Aufnahme zusätzlicher Mittel. Es wird dahingehend argumentiert, dass Unternehmen, die die Möglichkeit haben weitere finanzielle Mittel aufzutreiben, ihre Fähigkeit fortzubestehen erhöhen können, zumindest für den Zeitraum, den der Abschlussprüfer bei seiner Entscheidung über die Erteilung einer GCO betrachtet. Demgegenüber können für die Pläne zur Kürzung von Ausgaben bzw. Veräußerung von Vermögenswerten keine signifikanten Zusammenhänge zur Wahrscheinlichkeit der Erteilung einer GCO gefunden werden. Dennoch schließen Behn et al. (2001) aus den Befunden, dass Unternehmen zu diesen Plänen aufgrund von Liquiditätsengpässen gedrängt werden. Für Abschlussprüfer stellen Pläne zur Kürzung von Ausgaben bzw. Veräußerung von Vermögenswerten bei Unternehmen mit einer stärkeren (schwächeren) Liquiditätsposition ein positives (negatives) Zeichen dar.

Weiterhin wird die abschwächende Wirkung von kurzfristigem Liquiditätspotential (Short-Term Cash Potential) sowie strategischem Wachstum auf die Wahrscheinlichkeit der Erteilung einer GCO von Bruynseels und Willekens (2012) untersucht.[80] Für ihre Untersuchung ziehen die Autoren 389 Unternehmen mit einer GCO zwischen den Jahren 1998 bis 2001 heran und wählen hiervon zufällig 90 Unternehmen aus, die mit einer Kontrollgruppe aus Unternehmen in finanzieller Schieflage, aber ohne eine GCO, kombiniert werden. Die Ergebnisse weisen auf einen signifikant negativen Einfluss von strategischen Maßnahmen, die mit großer Wahrscheinlichkeit sowohl kurzfristig als auch langfristig einen positiven Cashflow erzielen, auf die Wahrscheinlichkeit des Erhalts einer GCO hin.[81] Im Gegensatz dazu wird die Ausstellung einer GCO bei Initiativen, die voraussichtlich nur auf lange Sicht einen positiven Cashflow erzielen, wahrscheinlicher. So werden bspw. Fusionen und Übernahmen nicht als mildernde Faktoren angesehen, da diese kurzfristig wahrscheinlich keine Verbesserung der finanziellen Lage herbeiführen und vielmehr einen weiteren Risikofaktor für die Erteilung einer GCO darstellen. Ebenso führen Kostenreduktionen, die einen positiven Einfluss auf die Liquidität haben sollen, eher zu einer höheren Wahrscheinlichkeit der Ausstellung einer GCO, da die kurzfristige Lösung nicht unbedingt langfristig gehalten werden kann.

Die Qualität der Finanzberichterstattung wird in Studien u. a. mithilfe der Accounting Accruals untersucht, die eine subjektive Einschätzung von zukünftigen Ergebnissen durch das Management darstellen und mit einer gewissen Unsicherheit verbunden sind.[82] Es wird angenommen, dass das Niveau der Accruals in Verbindung mit Problemen der Fortführung der Geschäftstätigkeit eines Unternehmens steht. J. R. Francis und Krishnan (1999) können bei ihrer Analyse von 2.608 Beobachtungen in den USA im Zeitraum von 1986 bis 1987 feststellen, dass bei Unternehmen mit einem hohen Anteil an Accruals die Erteilung einer GCO durch den Abschlussprüfer wahrscheinlicher ist.[83] Ebenso zeigt die Studie von Bartov, Gul und Tsui (2001) beim Vergleich von 346 Unternehmen, von denen die eine Hälfte einen eingeschränkten Bestätigungsvermerk und die andere Hälfte einen uneingeschränkten Bestätigungsvermerk aufweist, einen signifikanten Zusammenhang zwischen der Höhe der Discretionary Accruals und der Einschränkung des Prüfungsurteils.[84]

Dahingegen argumentieren Butler, Leone und Willenborg (2004), dass die Beurteilung der Qualität der Finanzberichterstattung nicht Aufgabe des Abschlussprüfers ist und ein Earnings Management, bspw. durch die Gestaltung der Abnormal Accruals, die Modifizierung des Bestätigungsvermerks nicht beeinflusst.[85] Für ihre Untersuchung wählen die Autoren die Periode von 1980 bis 1999 aus und stellen fest, dass bei rund 3.100 der über 7.000 betrachteten modifizierten Bestätigungsvermerken eine GCO zugrunde liegt. Belege für ein erhöhtes Earnings Management bei Unternehmen mit einem modifizierten Bestätigungsvermerk im Vergleich zu Unternehmen mit einem uneingeschränkten Prüfungsurteil können nicht gefunden werden. Die größere Wahrscheinlichkeit für die Erteilung einer GCO kann allerdings bei äußerst negativen Abnormal Accruals ermittelt werden, was auf die schlechte finanzielle Situation des Unternehmens zurückgeführt wird.

DeFond, Lim und Zang (2016) untersuchen den Zusammenhang zwischen dem Mandanten-Konservatismus und der Erteilung einer GCO im Zeitraum von 2000 bis 2010, wobei zur Ermittlung des Konservatismus neben der Querschnittsmethode (CON_KW) auch ein Modell mit Non-Operating Accruals (CON_NOPA) verwendet wird.[86] Es wird angenommen, dass konservative Mandanten ein geringeres Engagement-Risiko aufweisen und Abschlussprüfer daher weniger GCOs ausstellen. Die Ergebnisse zeigen einen signifikant negativen Koeffizienten für CON_KW sowie einen negativen, allerdings auf dem konventionellen Level nicht si-

[79] Vgl. hierfür und nachfolgend Behn et al. (2001, S. 17 u. 24-25).
[80] Vgl. hierfür und nachfolgend Bruynseels und Willekens (2012, S. 223 u. 230).
[81] Vgl. hierfür und nachfolgend Bruynseels und Willekens (2012, S. 236).
[82] Vgl. hierfür und nachfolgend J. R. Francis und Krishnan (1999, S. 135).
[83] Vgl. J. R. Francis und Krishnan (1999, S. 148 u. 157).
[84] Vgl. Bartov et al. (2001, S. 423-424).
[85] Vgl. hierfür und nachfolgend Butler et al. (2004, S. 140-142).
[86] Vgl. hierfür und nachfolgend DeFond et al. (2016, S. 70 u. 73).

gnifikanten, Koeffizienten für CON_NOPA.[87] Wenngleich die Ergebnisse für das zweite Modell schwächer sind, kann angenommen werden, dass Abschlussprüfer mit einer geringeren Wahrscheinlichkeit eine GCO für finanziell gestresste Unternehmen mit einer konservativen Finanzberichterstattung ausstellen. Die Erhöhung von CON_KW bzw. CON_NOPA um ein Zehntel verringert die Wahrscheinlichkeit einer GCO um 0,3 % bzw. 0,2 %.

Der Jahresabschluss eines Unternehmens stellt wichtige finanzielle Informationen für die Stakeholder bereit und eine hohe Qualität der Finanzberichterstattung kann die Informationsasymmetrien zwischen Unternehmen und Stakeholdern verringern.[88] Die Verbindung zwischen Informationsasymmetrien und einer Qualifizierung des Bestätigungsvermerks analysieren Abad et al. (2017) anhand spanischer Firmen in den Jahren von 2001 bis 2008.[89] Die Studie zeigt größere Informationsasymmetrien bei Unternehmen mit einem eingeschränkten Prüfungsurteil im Vergleich zu Unternehmen mit einem uneingeschränkten Bestätigungsvermerk, wobei die Differenzen insbesondere auf die Prüfungsurteile mit einer GCO zurückzuführen sind. Die Lesbarkeit ist ebenso ein Faktor, der sich auf die Qualität sowie den Nutzen der Finanzberichterstattung auswirkt und in der Untersuchung von *Abernathy et al.* (2019) aufgegriffen wird.[90] Die Autoren betrachten die Jahresabschlüsse von Unternehmen aus den Geschäftsjahren von 2001 bis 2014 und ermitteln die Lesbarkeit der Fußnoten anhand der Anzahl sowie dem Anteil an gängigen und finanziellen Wörtern. Während für die Anzahl der Wörter kein signifikanter Zusammenhang ermittelt werden kann, wird ein signifikant negativer Zusammenhang zwischen dem Anteil an gängigen sowie finanziellen Wörtern und der Wahrscheinlichkeit einer erstmaligen Ausstellung einer GCO durch den Abschlussprüfer nachgewiesen.[91] Für Unternehmen mit einer geringeren Lesbarkeit der Finanzberichterstattung ist der Erhalt einer erstmaligen GCO somit wahrscheinlicher als für Unternehmen mit einer höheren Lesbarkeit.

3.1.3. Corporate Governance und interne Kontrollen

Das im Rahmen des Corporate Governance eingesetzte Überwachungsorgan kann einen Einfluss auf die Ausstellung eines Bestätigungsvermerks mit einer GCO haben. Dies zeigen bspw. Carcello und Neal (2000) in ihrer Studie von 223 amerikanischen Unternehmen, die sich in einer finanziellen Schieflage befinden.[92] Die Untersuchung befasst sich mit dem Zusammenhang zwischen der Besetzung des Audit Committees mit abhängigen sowie unabhängigen Mitgliedern und der Berichterstattung des Abschlussprüfers. Die Autoren stellen fest, dass eine signifikant negative Verbindung zwischen dem Anteil an Affiliated Direktoren im Audit Committee und der Wahrscheinlichkeit der Ausstellung einer GCO besteht.[93] Je größer der Anteil an Affiliated Direktoren im Audit Committee ist, desto geringer ist die Wahrscheinlichkeit, eine GCO erteilt zu bekommen. Diese Erkenntnisse werden von Parker et al. (2005) gestützt, die eine Analyse von 161 finanziell gestressten Unternehmen im Zeitraum von 1988 bis 1996 durchführen.[94] Die Autoren können ebenfalls einen signifikanten Zusammenhang erkennen und ermitteln, dass ein abhängigeres Audit Committee, also ein Audit Committee mit einem höheren Anteil an angestellten Mitgliedern, mit einer geringeren Wahrscheinlichkeit eine GCO vom Abschlussprüfer erhält.

Seit dem Sarbanes-Oxley Act of 2002 (SOX) soll das Audit Committee ausnahmslos mit unabhängigen Direktoren, die keine finanziellen oder familiären Beziehungen zum Unternehmen oder dem CEO haben, besetzt sein.[95] Die Studie von Bruynseels und Cardinaels (2014) berücksichtigt diese Änderung und widmet sich der Beziehung zwischen den geschäftsführenden Direktoren und den Mitgliedern des Audit Committees.[96] Die Autoren analysieren Unternehmen in der Zeit von 2004 bis 2008 und unterscheiden in ihrer Untersuchung zwischen vollständig unabhängigen Mitgliedern des Audit Committees sowie solchen, die eine soziale Verbindung zum CEO vorweisen, bspw. hervorgerufen durch eine gemeinsame berufliche Tätigkeit, Ausbildung oder nicht-berufliche Aktivitäten. Sie kommen zu dem Schluss, dass die Qualität der Überwachung durch das Audit Committee und der Prüfungsaufwand aufgrund sozialer Verbindungen verringert ist, wodurch eine geringere Wahrscheinlichkeit der Ausstellung einer GCO resultiert. Hanlon, Khedmati und Lim (2019) betrachten ebenfalls die Beziehung zwischen dem CEO und den Direktoren bzw. den Direktoren des Audit Committees und weisen eine signifikant positive Verbindung zwischen Unternehmen nach, bei denen „Backscratching“ zu finden und die Wahrscheinlichkeit, eine GCO zu erhalten, hoch ist.[97] Bei Unternehmen mit „Backscratching“ erhalten sowohl der CEO als auch die Direktoren gleichzeitig eine Vergütung, die oberhalb der des Marktes liegt. Wie die Analyse zeigt, ist diese Eigenschaft mit einer schlechteren zukünftigen Leistung des Unternehmens und einer erhöhten Wahrscheinlichkeit des Erhalts einer GCO verbunden.

Die Wirksamkeit der internen Kontrolle hat ebenso einen Einfluss auf die Wahrscheinlichkeit der Ausstellung einer GCO, wie Goh et al. (2013) in ihrer Untersuchung von 1.100 Unternehmen, die in der Periode von 2004 bis 2009 eine finanziell angespannte Situation aufweisen, feststellen.[98] Es kann nachgewiesen werden, dass bei einer Erteilung eines Prüfungsurteils über wesentliche Schwachstellen in der internen Kontrolle (Material Weakness Opinion), die Wahrscheinlichkeit für die Ausstellung einer GCO signifikant

[87] Vgl. hierfür und nachfolgend DeFond et al. (2016, S. 75).
[88] Vgl. Abad, Sánchez-Ballesta und Yagüe (2017, S. 565).
[89] Vgl. hierfür und nachfolgend Abad et al. (2017, S. 589).
[90] Vgl. hierfür und nachfolgend Abernathy, Guo, Kubick und Masli (2019, S. 6).
[91] Vgl. hierfür und nachfolgend Abernathy et al. (2019, S. 12).
[92] Vgl. hierfür und nachfolgend Carcello und Neal (2000, S. 453 u. 457).
[93] Vgl. hierfür und nachfolgend Carcello und Neal (2000, S. 465).
[94] Vgl. hierfür und nachfolgend Parker et al. (2005, S. 11 u. 18).
[95] Vgl. Bruynseels und Cardinaels (2014, S. 116).
[96] Vgl. hierfür und nachfolgend Bruynseels und Cardinaels (2014, S. 114-115).
[97] Vgl. hierfür und nachfolgend Hanlon et al. (2019, S. 198-199).
[98] Vgl. hierfür und nachfolgend Goh et al. (2013, S. 972).

höher ist. Gramling, Krishnan und Zhang (2011) unterstützen die Erkenntnis durch die Untersuchung von finanziell angeschlagenen Unternehmen in den Jahren von 2004 bis 2006, die eine positive Verbindung zwischen einer wesentlichen Schwäche in Bezug auf den Prüfungsausschuss und der Wahrscheinlichkeit des Erhalts einer GCO aufzeigt.[99] Eine Schwäche des Audit Committees liegt nach Auslegung der Studie dann vor, wenn das Unternehmen in seinem vierteljährlichen Bericht nach SOX Section 302 über eine interne Kontrollschwäche des Audit Committees berichtet. Ebenso kommen Jiang, Rupley und Wu (2010) zu dem gleichen Befund und diagnostizieren eine mehr als dreifach so hohe Wahrscheinlichkeit für den Erhalt einer GCO, wenn bei einem Unternehmen wesentliche Schwächen in der internen Kontrolle vorliegen.[100] Sie begründen dies damit, dass Unternehmen mit Mängeln in der internen Überwachung aufgrund der höheren Wahrscheinlichkeit an unentdeckten Going-Concern-Themen und Falschangaben ein größeres Risiko aufweisen.[101] So könnten bspw. Hinweise auf Probleme bei der Unternehmensfortführung durch das Management einfacher verborgen werden.

Dass eine funktionierende und effektive interne Kontrolle die Wahrscheinlichkeit der Erteilung einer GCO beeinflussen kann, zeigt sich darüber hinaus in den Ergebnissen von Hassan und Nasir (2020).[102] Sie betrachten finanziell angeschlagene Unternehmen in Malaysia und erkennen einen signifikant negativen Zusammenhang zwischen den Investitionen in die interne Revision und der Going-Concern-Berichterstattung des externen Abschlussprüfers. Die Ausgaben für die interne Revision führen zu effektiven und effizienten Kontrollen, die wiederum die Wahrscheinlichkeit für den Erhalt einer GCO verringern.

3.2. Faktoren beim Abschlussprüfer

3.2.1. Eigenschaften des Abschlussprüfers

Die Eigenschaften und Sichtweisen des Abschlussprüfers sind bei der Entscheidung über die Erteilung einer GCO von entscheidender Bedeutung, wie bspw. die frühe Studie von Kida (1980) zeigt.[103] Abschlussprüfer, die weniger eingeschränkte Bestätigungsvermerke ausstellen, sind stärker davon überzeugt, dass sie durch eine Einschränkung des Bestätigungsvermerks den Mandanten verlieren oder von ihm verklagt werden und die Reputation der Prüfungsgesellschaft negativ beeinflusst wird.

Des Weiteren wird die Urteilsbildung des Abschlussprüfers durch seine Ausbildung und Erfahrung beeinflusst. In ihrem Experiment teilen Lehmann und Norman (2006) die Teilnehmer in zwei Gruppen ein, auf der einen Seite die Neuanfänger und auf der anderen Seite die Fachleute.[104] Alle Teilnehmer erhalten einen Fall, der auf echten Daten basiert, wobei das Unternehmen verschiedene Probleme aufweist. Die Aufgabe der Teilnehmer ist es, die finanziellen Rahmenbedingungen des Unternehmens zusammenzufassen und die Art eines Bestätigungsvermerks zu bestimmen. Die Autoren kommen in ihrer Untersuchung zu dem Ergebnis, dass erfahrenere Abschlussprüfer im Vergleich zu Anfängern die Probleme übersichtlicher und präziser darstellen.[105] Im Kontrast dazu ist kein wesentlicher Unterschied zwischen den Gruppen hinsichtlich der Ausstellung einer GCO erkennbar, weil die Mehrheit der Teilnehmer die Lage des Unternehmens richtig einschätzt. Die Genauigkeit der Going-Concern-Berichte von Unternehmen in Norwegen ist Gegenstand der Studie von Che, Langli und Svanström (2018), die über 170.000 Beobachtungen von Geschäftsjahren von Mandanten auswertet, die von über 1.700 verschiedenen Prüfungspartnern geprüft wurden.[106] Die Autoren kommen zu dem Schluss, dass die Genauigkeit der Going-Concern-Berichte bei Abschlussprüfern mit einem Masterabschluss im Vergleich zu Abschlussprüfern mit einem Bachelorabschluss sowie bei Prüfern mit einer größeren Anzahl an beruflichen Fortbildungskursen und der meisten Erfahrung, gemessen anhand der Anzahl der Jahre seit Erhalt der ersten Prüferlizenz, höher ist.[107]

Dagegen stellen Hardies, Breesch und Branson (2016) bei der Auswertung ihrer Stichprobe von 7.105 finanziell angeschlagenen privaten belgischen Unternehmen fest, dass eine signifikant negative Verbindung zwischen der Erfahrung des Prüfungspartners, gemessen an der Anzahl der Jahre, seitdem der Abschlussprüfer den Bestätigungsvermerk unterzeichnen darf, und der Wahrscheinlichkeit des Erhalts einer GCO besteht.[108] Hat der Prüfungspartner eine größere Erfahrung, dann ist es für ein Unternehmen weniger wahrscheinlich eine GCO zu erhalten. Eine andere Schlussfolgerung ergibt sich jedoch aus der Studie von Hardies, Vandenhaute und Breesch (2018).[109] Hier wird die Genauigkeit der Berichterstattung über die Unternehmensfortführung bei privat gehaltenen Unternehmen in Belgien analysiert und die Autoren betrachten u. a. die falsch negativen Klassifizierungen, d. h. der Abschlussprüfer stellt einem Unternehmen, das innerhalb des nächsten Jahres seine Tätigkeit einstellt, keine GCO aus. Diese Fehleinschätzung ist bekannt als Typ II Fehler bzw. Type II Misclassification. Im Gegensatz dazu wird von einem Typ I Fehler gesprochen, wenn der Abschlussprüfer eine GCO erteilt, das Unternehmen anschließend jedoch nicht aufhört zu existieren und weiterhin besteht. In den Ergebnissen von Hardies et al. (2018) lassen sich im Vergleich zu der früheren Studie keine signifikanten Hinweise dafür finden, dass erfahrenere Prüfungspartner im Vergleich zu weniger erfahrenen Prüfungspartnern eine geringere Rate an falsch negativen Fehlklassifizierungen aufweisen.

Zur Untersuchung des Einflusses der Erfahrung des Ab-

[99] Vgl. hierfür und nachfolgend Gramling et al. (2011, S. 63 u. 69-70).
[100] Vgl. Jiang et al. (2010, S. 43).
[101] Vgl. hierfür und nachfolgend Jiang et al. (2010, S. 40).
[102] Vgl. hierfür und nachfolgend Hassan und Nasir (2020, S. 34).
[103] Vgl. hierfür und nachfolgend Kida (1980, S. 516).
[104] Vgl. hierfür und nachfolgend Lehmann und Norman (2006, S. 71-72).
[105] Vgl. hierfür und nachfolgend Lehmann und Norman (2006, S. 79-80).
[106] Vgl. Che et al. (2018, S. 91).
[107] Vgl. Che et al. (2018, S. 105).
[108] Vgl. hierfür und nachfolgend Hardies et al. (2016, S. 19).
[109] Vgl. hierfür und nachfolgend Hardies et al. (2018, S. 119 u. 129).

schlussprüfers auf die Erteilung einer GCO, kann die Erfahrung durch das Alter des Abschlussprüfers dargestellt werden.[110] Sundgren und Svanström (2014) verwenden das Alter des Abschlussprüfers in ihrer Studie schwedischer Unternehmen, die im betrachteten Zeitraum einen Insolvenzantrag gestellt haben, und stellen einen Bezug zur Erteilung einer GCO her. Die Analyse zeigt, dass ältere Abschlussprüfer eine geringere Wahrscheinlichkeit aufweisen, eine GCO auszustellen, als jüngere Abschlussprüfer. Eine ähnliche Verbindung erkennen Goodwin und Wu (2016) in ihrer Untersuchung.[111] Bei der Auswertung ihrer Stichprobe, die aus gelisteten australischen Unternehmen im Zeitraum von 1999 bis 2010 besteht, ist erkennbar, dass ältere Prüfungspartner mit einer geringeren Wahrscheinlichkeit eine GCO erteilen. Ebenso zeigen die Ergebnisse bei älteren Prüfungspartnern eine geringere Genauigkeit bei der Ausstellung einer GCO innerhalb eines Zweijahreszeitraumes. Mögliche Erklärungsansätze für den negativen Zusammenhang sind ein geringeres Bemühen für das richtige Verstehen und Umsetzen von Prüfungsstandards, eine Entscheidung zur Nichtberücksichtigung der Informationen oder eine größere Nachsichtigkeit durch ältere Abschlussprüfer.[112]

Ein weiterer Aspekt ist die Spezialisierung des Abschlussprüfers in einer Industrie, die in der Forschung zu unterschiedlichen Ansichten führt. Reichelt und Wang (2010) unterscheiden in ihrer Untersuchung zwischen der Industriespezialisierung auf nationaler und lokaler Ebene.[113] Darüber hinaus werden zwei Definitionen für die Spezialisierung verwendet. Die erste Definition beschreibt Industriespezialisten, wenn der Abschlussprüfer den größten Marktanteil in einer zweistelligen SIC-Kategorie hat und dieser mindestens 10 % größer ist als der der folgenden Abschlussprüfer. Dagegen ist nach der zweiten Definition ein Abschlussprüfer ein Experte in der Industrie, wenn er einen Marktanteil von mindestens 30 % in einer zweistelligen SIC-Kategorie verzeichnen kann. Die Autoren gehen davon aus, dass ein Abschlussprüfer, der in einer Industrie spezialisiert ist, strenger zu seinen Mandanten ist und er daher mit einer größeren Wahrscheinlichkeit eine GCO erteilt.[114] Es wird festgestellt, dass eine rund 10 % höhere Wahrscheinlichkeit besteht, eine GCO zu erhalten, wenn es sich um einen sowohl auf nationaler als auch auf lokaler Ebene industriespezialisierten Abschlussprüfer handelt. Werden die Ebenen einzeln betrachtet, so ist bei beiden Definitionen ein positiver Zusammenhang nachweisbar, wobei dieser nur auf lokaler Ebene nach der zweiten Definition signifikant ist. Bills, Jeter und Stein (2015) erkennen eine positive und signifikante Verbindung zwischen der Prüfung durch einen spezialisierten Abschlussprüfer und der Wahrscheinlichkeit der Erteilung einer GCO.[115] Die Signifikanz des Effektes entfällt, wenn die Variablen für homogene und komplexe Industrien in die Analyse mit einbezogen werden. Das bedeutet, dass sich die Wahrscheinlichkeit für die Ausstellung einer GCO durch einen in einer homogenen bzw. homogenen und komplexen Industrie spezialisierten Abschlussprüfer nicht von der eines in einer anderen Industrie spezialisierten Abschlussprüfers unterscheidet.

Die Studie von Che et al. (2018) betrachtet hingegen die Industrieerfahrung auf Partnerebene, wobei diese anhand der aggregierten Umsätze aller Mandanten eines jeden Partners geteilt durch den Gesamtumsatz in der Branche gemessen wird.[116] Die Ergebnisse deuten darauf hin, dass die Genauigkeit der GCO bei Abschlussprüfern, die eine größere Erfahrung in der Industrie besitzen, höher ist und somit eine größere Prüfungsqualität gegeben ist.[117]

Dagegen können Sundgren und Svanström (2014) in ihrer Analyse von schwedischen Unternehmen in den Jahren 2006 und 2009 keine signifikanten Auswirkungen der Spezialisierung des Abschlussprüfers auf die Wahrscheinlichkeit einer GCO entdecken.[118] Sowohl auf Ebene der Prüfungsgesellschaft als auch auf Ebene des Büros und der des verantwortlichen Abschlussprüfers sind die Variablen in der Regression insignifikant. Ebenso führen die Unternehmensbeobachtungen durch Gaver und Utke (2019) in den Jahren von 2003 bis 2015 zu keinen Belegen, dass die Spezialisierung des Abschlussprüfers einen Einfluss auf die Erteilung einer GCO und die Prüfungsqualität hat.[119]

3.2.2. Finanzielle Abhängigkeit

Die finanzielle Abhängigkeit des Abschlussprüfers von seinen Mandanten ist ein Thema, welches sowohl in der Literatur als auch von den gesetzgebenden Institutionen intensiv diskutiert wird.[120] Der Meinungsaustausch basiert dabei auf der Annahme, dass der Abschlussprüfer gewillt ist, seine Unabhängigkeit in der Urteilsbildung zum Erhalt seiner Mandanten, die hohe Honorare für Prüfungs- sowie Nichtprüfungsdienstleistungen zahlen, einzuschränken. Gleichzeitig gilt es zu berücksichtigen, dass insbesondere mögliche Haftungsrisiken und Verfahrenskosten sowie der Reputationsverlust, die aus einer unzureichenden Unabhängigkeit resultieren können, eine Motivation zur Aufrechterhaltung der Unabhängigkeit des Abschlussprüfers darstellen können.

Für den Zeitraum vor und direkt nach dem SOX kommen Untersuchungen zu dem Ergebnis, dass in den USA kein signifikanter Zusammenhang zwischen dem Umfang an Nichtprüfungshonoraren, sog. Non-Audit Service (NAS) Fees, und der Wahrscheinlichkeit einer GCO im Bestätigungsvermerk besteht.[121] Ebenfalls können Callaghan, Parkash und Singhal (2009) bei ihrer Analyse von Unternehmen, die in der Zeit von Januar 2001 bis März 2005 Insolvenz angemeldet haben, keine Belege für einen signifikant negativen Zusammenhang

[110] Vgl. hierfür und nachfolgend Sundgren und Svanström (2014, S. 531 u. 543).
[111] Vgl. hierfür und nachfolgend Goodwin und Wu (2016, S. 363-364).
[112] Vgl. Sundgren und Svanström (2014, S. 549).
[113] Vgl. hierfür und nachfolgend Reichelt und Wang (2010, S. 656).
[114] Vgl. hierfür und nachfolgend Reichelt und Wang (2010, S. 676–677).
[115] Vgl. hierfür und nachfolgend Bills et al. (2015, S. 1746).
[116] Vgl. Che et al. (2018, S. 98).
[117] Vgl. Che et al. (2018, S. 105).
[118] Vgl. hierfür und nachfolgend Sundgren und Svanström (2014, S. 547).
[119] Vgl. Gaver und Utke (2019, S. 139).
[120] Vgl. hierfür und nachfolgend DeFond et al. (2002, S. 1248).
[121] Vgl. DeFond et al. (2002, S. 1271); Geiger und Rama (2003, S. 65).

zwischen den Nichtprüfungshonoraren und der Tendenz zu einer Erteilung einer GCO durch den Abschlussprüfer feststellen.[122] DeFond et al. (2002) erläutern, dass die Ergebnisse auf eine höhere Gewichtung der institutionellen Anreize durch den Abschlussprüfer hindeuten, wie bspw. den Verlust der Reputation und die möglichen Verfahrenskosten, als dass eine wirtschaftliche Abhängigkeit durch höhere Honorare von größerer Bedeutung ist.[123]

Im Gegensatz zu den bisher dargestellten Erkenntnissen ergibt sich bei Blay und Geiger (2013), die Unternehmen analysieren, welche in den Jahren von 2004 bis 2006 erstmalig eine GCO erhalten haben, ein anderes Ergebnis.[124] Die Autoren können in ihrer Studie eine signifikant negative Verbindung zwischen den aktuellen Nichtprüfungshonoraren und dem Erhalt einer GCO identifizieren, womit die Wahrscheinlichkeit für die Ausstellung einer GCO signifikant geringer ist, wenn die aktuellen Nichtprüfungshonorare hoch sind. Unterstützung erhalten diese Befunde durch die Ergebnisse der Studie von Griffin und Lont (2010), die 7.093 Beobachtungen von finanziell gestressten Unternehmen über die Jahre von 2000 bis 2006 auswertet, wobei hiervon 310 Unternehmen eine erstmalige GCO erhalten haben.[125] Es kann ebenfalls ein negativer Zusammenhang zwischen den Nichtprüfungshonoraren und der Neigung des Abschlussprüfers, eine GCO auszustellen, festgehalten werden.[126] Jedoch ist die negative Verbindung nur für die Zeit vor 2004 signifikant, sodass der Zusammenhang im post-2003-Zeitraum geringer ist. Die Autoren schließen daraus, dass die regulatorischen Änderungen durch den SOX eine effektive Wirkung zeigen und die Unabhängigkeit des Abschlussprüfers gefestigt wird.

Dahingegen wird von Robinson (2008) bei ihrer Untersuchung von insolventen Unternehmen in der Zeit von 2001 bis 2004 eine positive Korrelation zwischen dem Umfang an Honoraren für Steuerberatungsdienstleistungen und der Wahrscheinlichkeit einer korrekt ausgestellten GCO erkannt.[127] Sie argumentiert, dass die Unabhängigkeit des Abschlussprüfers durch diese Nichtprüfungsdienstleistung nicht beeinträchtigt wird, sondern die Prüfungsqualität vielmehr durch den Information-Spillover-Effekt, also der Gewinnung zusätzlicher Informationen im Zusammenhang mit der erbrachten Dienstleistung, zunimmt. Allerdings kann sie weder für die Prüfungshonorare noch für die Nichtprüfungshonorare (ohne Berücksichtigung der Komponenten für Steuerberatungsdienstleistungen) eine signifikante Verbindung zur Erteilung einer GCO feststellen.

Ergebnisse, die außerhalb der USA nachgewiesen werden, führen ebenfalls zu gemischten Erkenntnissen. Für Deutschland kann Ratzinger-Sakel (2013) in ihrer Analyse finanziell angeschlagener Unternehmen des produzierenden Gewerbes im Zeitraum von 2005 bis 2009 keine Hinweise dafür erkennen, dass deutsche Abschlussprüfer weniger unabhängig sind, wenn das Niveau der Nichtprüfungshonorare hoch ist.[128] Allerdings gibt es schwache Belege dafür, dass Big-4-Prüfungsgesellschaften[129] im Vergleich zu Non-Big 4-Prüfungsgesellschaften mit einer geringeren Wahrscheinlichkeit eine GCO erteilen, wenn der Mandant einen relativ großen Umfang an Nichtprüfungshonoraren zahlt und sich in einer extremen finanziellen Notlage befindet. Dagegen ergeben sich im Vereinigten Königreich (UK) zunächst keine Anhaltspunkte für eine Verbindung zwischen den Nichtprüfungshonoraren und der Wahrscheinlichkeit, eine GCO zu erhalten, wie die Studien mit finanziell gestressten Unternehmen von Lennox (1999) in der Zeit von 1988 bis 1994 und Wu, Hsu und Haslam (2016) im Zeitraum von 1997 bis 2010 zeigen.[130] Wu et al. (2016) stellen allerdings fest, dass die Wechselwirkungen zwischen den Nichtprüfungshonoraren und dem Anteil an unabhängigen Mitgliedern sowie finanziellen Experten im Prüfungsausschuss in einem positiven Zusammenhang zu der Wahrscheinlichkeit der Ausstellung einer GCO stehen.[131] Diese Beobachtung ist jedoch nur gültig, wenn der Anteil an unabhängigen Mitgliedern bzw. finanziellen Experten gering ist, da bei einem höheren Anteil ein insignifikantes Ergebnis ermittelt wird. Demgegenüber findet Firth (2002) Belege für einen signifikant negativen Zusammenhang zwischen dem Umfang an Beratungshonoraren und der Wahrscheinlichkeit einer GCO bei Unternehmen mit einem Ende des Geschäftsjahres im Jahr 1986.[132] Ebenso ermitteln Basioudis, Papakonstantinou und Geiger (2008) bei ihrer Auswertung von Unternehmen aus dem UK mit einem Bilanzstichtag im Jahr 2003 eine signifikant negative Verbindung zwischen der Größenordnung der Nichtprüfungshonorare und dem Erhalt einer GCO durch den Abschlussprüfer.[133]

Weiterhin sind die Prüfungshonorare selbst Gegenstand wissenschaftlicher Untersuchungen. Während Geiger und Rama (2003) eine positive Relation zwischen den Prüfungshonoraren und der Wahrscheinlichkeit der Ausstellung einer GCO erkennen, ergeben die Analysen von DeFond et al. (2002) und Blay und Geiger (2013) keinen Zusammenhang zwischen den Prüfungshonoraren und der Wahrscheinlichkeit der Erteilung einer GCO durch den Abschlussprüfer.[134] Allerdings finden auch Callaghan et al. (2009) keine Verbindung zwischen den Prüfungshonoraren und der Wahrscheinlichkeit des Erhalts einer GCO.[135]

Darüber hinaus wird der Einfluss der Größe des Man-

[122] Vgl. Callaghan et al. (2009, S. 166).
[123] Vgl. DeFond et al. (2002, S. 1271-1272).
[124] Vgl. hierfür und nachfolgend Blay und Geiger (2013, S. 585 u. 598).
[125] Vgl. Griffin und Lont (2010, S. 15).
[126] Vgl. hierfür und nachfolgend Griffin und Lont (2010, S. 20 u. 30).
[127] Vgl. hierfür und nachfolgend Robinson (2008, S. 34 u. 51-52).
[128] Vgl. hierfür und nachfolgend Ratzinger-Sakel (2013, S. 154-155).
[129] Bei den Big-4-Prüfungsgesellschaften handelt es sich um die vier größten und umsatzstärksten Wirtschaftsprüfungsgesellschaften (Deloitte, EY, KPMG, PwC). Die Big-4-Prüfungsgesellschaften haben sich über die Zeit aus Fusionierungen der damals acht größten Wirtschaftsprüfungsgesellschaften entwickelt.
[130] Vgl. Lennox (1999, S. 248); Wu et al. (2016, S. 253-254).
[131] Vgl. hierfür und nachfolgend Wu et al. (2016, S. 253-254).
[132] Vgl. Firth (2002, S. 683).
[133] Vgl. Basioudis et al. (2008, S. 300).
[134] Vgl. Blay und Geiger (2013, S. 598); DeFond et al. (2002, S. 1271); Geiger und Rama (2003, S. 66).
[135] Vgl. Callaghan et al. (2009, S. 166).

danten auf die ökonomische Abhängigkeit des Abschlussprüfers diskutiert, da ein hoher Anteil der Prüfungshonorare eines Mandanten an den gesamten Umsätzen einer örtlichen Niederlassung einer Prüfungsgesellschaft im Vergleich zu einem niedrigeren Anteil die Entscheidungen des Abschlussprüfers beeinflussen könnte.[136] Reynolds und Francis (2000) finden in der Untersuchung ihrer Stichprobe keine Beweise für einen signifikanten Effekt auf die Entscheidung des Abschlussprüfers über die GCO in Abhängigkeit von der Größe des Mandanten im Verhältnis zur Größe des Standortes einer Prüfungsgesellschaft. Ebenfalls zu keinem statistisch signifikanten Zusammenhang kommt Li (2009) bei ihrer Studie über die Wichtigkeit eines Mandanten, gemessen am Anteil des Umsatzes eines Mandanten an den gesamten Umsätzen einer örtlichen Niederlassung, wobei die Umsätze sowohl als gesamtes Honorar als auch getrennt in Prüfungshonorare und Nichtprüfungshonorare untersucht werden.[137] Diese Erkenntnis ist allerdings nur für das Jahr 2001 (pre-SOX) gültig, da für das Jahr 2003 (post-SOX) bei den Prüfungshonoraren und dem gesamten Honorar ein positiver Zusammenhang zur GCO ermittelt wurde. Kao, Li und Zhang (2014) überprüfen die Ergebnisse von Li (2009) in ihrer Analyse und können dabei die Erkenntnisse über die Honorarabhängigkeit, definiert als Verhältnis der gesamten Prüfungs- und Nichtprüfungshonorare eines Mandanten zum gesamten Umsatz einer örtlichen Niederlassung des Abschlussprüfers, in den Jahren 2001 und 2003 bestätigen.[138] Dahingegen wird bei einer Ausweitung der Untersuchung über die Jahre von 2004 bis 2009 keine Korrelation der Honorarabhängigkeit mit der Ausstellung einer GCO gefunden.

3.2.3. Größe des Abschlussprüfers

Die Größe der Wirtschaftsprüfungsgesellschaft ist ebenfalls ein Indikator, der in der Literatur als ein möglicher Einflussfaktor auf die Erteilung einer GCO durch den Abschlussprüfer angesprochen wird. Hierbei wird zwischen den größten Prüfungsgesellschaften, den Big-4, sowie den nachfolgenden, kleineren Prüfungsgesellschaften, den Non-Big-4, unterschieden. Die Studie von Mutchler et al. (1997) betrachtet Unternehmen in Insolvenz im Zeitraum von 1974 bis 1985 und untersucht den Einfluss der damaligen Big-6- und Non-Big-6-Prüfungsgesellschaften auf die Wahrscheinlichkeit der Ausstellung einer GCO.[139] Die Autoren können dabei keinen signifikanten Unterschied zwischen den Big-6- und den Non-Big-6-Prüfungsgesellschaften feststellen.

Dahingegen wird sowohl von Reichelt und Wang (2010) als auch Numan und Willekens (2012) eine negative Verbindung ermittelt, d. h. Mandanten von Big-4-Prüfungsgesellschaften haben eine signifikant geringere Wahrscheinlichkeit, eine GCO zu erhalten.[140] Der negative Zusammenhang wird damit begründet, dass die Korrelation aus der besseren finanziellen Situation der durch die Big-4 geprüften Mandanten resultiert und für die Unternehmen daher weniger die Notwendigkeit einer GCO besteht.[141] Bei der Untersuchung australischer Bestätigungsvermerke von 2005 bis 2013 sind für Carson, Fargher und Zhang (2016) bei Non-Big-4-Prüfungsgesellschaft eine signifikant höhere Anzahl an GCOs erkennbar als im Vergleich zu Big-4-Prüfungsgesellschaften.[142] Die Ergebnisse folgen der Ansicht, dass die Mandanten der Big-4 eine durchschnittlich größere Unternehmensgröße und ein niedrigeres Niveau von Risiken aufweisen, wodurch der Abschlussprüfer mit einer geringeren Wahrscheinlichkeit dazu geneigt ist, eine GCO auszustellen. Zudem werden mögliche Differenzen hinsichtlich rechtlicher Bedenken zwischen den Non-Big-4 und den Big-4 als denkbare Gründe für den Unterschied in der Wahrscheinlichkeit der Erteilung einer GCO angeführt.

Habib (2013) schließt dagegen aus den Ergebnissen seiner Meta-Studie, dass ein positiver Zusammenhang zwischen der Größe der Prüfungsgesellschaft und der Wahrscheinlichkeit der Erteilung einer GCO vorhanden ist und somit die Ausstellung einer GCO durch eine Big-4-Prüfungsgesellschaft wahrscheinlicher ist als durch eine Non-Big-4-Prüfungsgesellschaft.[143] Die Feststellung kann nach Ansicht des Autors mit der Lieferung einer qualitativ hochwertigen Prüfung und der Sicherung des Markenwerts, in den signifikante Ressourcen investiert wurden, begründet werden. Dieser Argumentation folgen J. R. Francis und Yu (2009), die bei der Auswertung von Bestätigungsvermerken mit einer GCO in der Periode von 2003 bis 2005 eine signifikant positive Verbindung zwischen der Größe des Büros des Abschlussprüfers und der Wahrscheinlichkeit des Erhalts einer GCO dokumentieren.[144] Sie begründen ihre Befunde damit, dass größere Büros eine höhere Genauigkeit in der Berichterstattung der GCO vorweisen können und die Qualität der Prüfung größer ist.

Die unterschiedlichen Richtungen der Zusammenhänge in den Forschungsergebnissen ergeben sich nach Berglund et al. (2018) aus den verwendeten Untersuchungsdesigns.[145] Demnach würden die Studien nicht angemessen die finanzielle Schieflage des Mandanten kontrollieren und die Verwendung von Variablen, die die Eigenschaften des Abschlussprüfers darstellen, führe zu einer Verzerrung der Testergebnisse. Daher versuchen die Autoren in ihrer Stichprobe, die sich über den Zeitraum von 2000 bis 2013 erstreckt, sowie in ihrem Modell die erörterten Probleme zu vermeiden. Als Resultat ergibt sich eine positive Verbindung zwischen der Größe des Abschlussprüfers und der Tendenz der Ausstellung einer GCO, d. h. Mandanten von Big-4-Prüfungsgesellschaften haben eine signifikant höhere Wahrscheinlichkeit, eine GCO zu erhalten, als Mandanten von Non-Big-4-Prüfungsgesellschaften. Der positive Zusammenhang kann damit begründet werden, dass sich Big-4-

[136] Vgl. hierfür und nachfolgend Reynolds und Francis (2000, S. 376-377).
[137] Vgl. hierfür und nachfolgend Li (2009, S. 225).
[138] Vgl. hierfür und nachfolgend Kao et al. (2014, S. 183).
[139] Vgl. hierfür und nachfolgend Mutchler et al. (1997, S. 302-304).
[140] Vgl. Numan und Willekens (2012, S. 23); Reichelt und Wang (2010, S. 677).
[141] Vgl. Carson et al. (2013, S. 361).
[142] Vgl. hierfür und nachfolgend Carson et al. (2016, S. 234-235).
[143] Vgl. hierfür und nachfolgend Habib (2013, S. 200-203).
[144] Vgl. hierfür und nachfolgend J. R. Francis und Yu (2009, S. 1545-1546).
[145] Vgl. hierfür und nachfolgend Berglund et al. (2018, S. 19-20).

Prüfungsgesellschaften bei einem Typ II Fehler höheren Kosten für Reputationsschäden und Rechtsstreitigkeiten ausgesetzt sehen und daher konservativer berichten.[146] Die Folgekosten eines Typ I Fehlers, hierzu gehört der Verlust eines Mandanten, sind für größere Prüfungsgesellschaften aufgrund der größeren Mandantenbasis weniger relevant, weshalb die Berichterstattung bei Big-4-Prüfungsgesellschaften weniger konservativ ist.

Die Größe des Abschlussprüfers steht darüber hinaus mit weiteren Faktoren in Verbindung. So werden bspw. von Mo, Rui und Wu (2015) Belege dafür gefunden, dass die Wirkung von Gesetzen, die die GCO betreffen, von der Größe des Abschlussprüfers abhängig ist.[147] Bei der Auswertung der Going-Concern-Berichterstattung vor und nach dem ersten chinesischen Insolvenzgesetz im August 2006 wird eine höhere Tendenz zur Ausstellung einer GCO bei den chinesischen Big-4-Prüfungsgesellschaften in der Zeit vor dem Insolvenzgesetz im Vergleich zu den lokalen Top-10-Abschlussprüfern ermittelt, wohingegen dieser Unterschied nach Veröffentlichung des Gesetzes nicht mehr vorhanden ist. Beim Vergleich von Nicht-Top-10-Abschlussprüfern mit den Big-4-Prüfungsgesellschaften wird sowohl in der ante- als auch post-Phase eine höhere Wahrscheinlichkeit der Ausstellung einer GCO bei den chinesischen Big-4-Prüfungsgesellschaften festgestellt. Diese Befunde deuten nach Meinung der Autoren darauf hin, dass bei einem unterentwickelten rechtlichen Umfeld die chinesischen Big-4-Prüfungsgesellschaften aufgrund der internationalen Reputation bereits eine strengere Beurteilung bei der Frage der Unternehmensfortführung verfolgen. Die Top-10-Abschlussprüfer reagieren dagegen auf die Rechtsstreitigkeits- und Haftungsrisiken und ändern ihr Berichtsverhalten.

Auch bei der Mandatsdauer der Prüfungsgesellschaft und einem Prüfungsversagen in Form eines Typ II Fehlers, also das Versäumnis der Ausstellung einer GCO, gibt es Unterschiede zwischen den Big-4- und den Non-Big-4-Prüfungsgesellschaften, wie Read und Yezegel (2016) zeigen.[148] Den Ergebnissen nach ist bei den Big-4-Prüfungsgesellschaften kein Zusammenhang zwischen der Mandatsdauer und einem Prüfungsversagen erkennbar. Im Gegensatz dazu lässt sich bei den Non-Big-4-Prüfungsgesellschaften eine Relation ermitteln, die nicht linear verläuft. Während der ersten Jahre einer Beschäftigung stellen Non-Big-4-Prüfungsgesellschaften mit einer geringeren Wahrscheinlichkeit eine GCO zu einem anschließend insolventen Mandanten aus. In den folgenden Jahren, also bei einer langen Mandatsdauer, ist kein Effekt auf einen Fehler in der Prüfung vorhanden.

3.3. Beziehung zwischen Abschlussprüfer und Unternehmen

3.3.1. Beschäftigungsdauer des Abschlussprüfers

Bereits seit längerer Zeit wird sowohl von Wissenschaftlern als auch Regulierungsbehörden diskutiert, dass die Unabhängigkeit und Objektivität des Abschlussprüfers durch eine längere Beschäftigungsdauer bei einem Mandanten negativ beeinflusst wird, weshalb die Rotation des Abschlussprüfers zu einer neuen Perspektive und einer geringeren Beeinflussung des Abschlussprüfers durch den Mandanten und einer damit einhergehenden Steigerung der Prüfungsqualität führen soll.[149] Ob eine längere Mandatsdauer tatsächlich einen Einfluss auf die Prüfungsqualität hat, wird in vielen Studien mithilfe der Going-Concern-Berichterstattung untersucht.

Die angeführte Vermutung wird von Carey und Simnett (2006) belegt, die bei australischen Unternehmen im Jahr 1995 feststellen, dass die Tendenz der Ausstellung einer GCO über die Mandatsdauer des Prüfungspartners abnimmt.[150] Diese Erkenntnis wird unterstützt von Ye et al. (2011), die ebenfalls Unternehmen in Australien untersuchen, allerdings für das Jahr 2002.[151] Sie ermitteln bei einer längeren Beschäftigungsdauer des Prüfungspartners zwar eine Abnahme in der Neigung des Abschlussprüfers eine angemessene GCO zu erteilen. Einen Zusammenhang zwischen der Beschäftigungsdauer der Prüfungsgesellschaft und einer GCO können sie allerdings nicht feststellen.

Im Gegensatz zu der Annahme einer negativen Beziehung zwischen der Beschäftigungsdauer und der Wahrscheinlichkeit des Erhalts einer GCO, ermitteln Geiger und Raghunandan (2002) in ihrer Untersuchung von Unternehmen in den USA, die in den Jahren von 1996 bis 1998 einen Insolvenzantrag gestellt haben, eine positive Verbindung.[152] Den Befunden nach ist es somit bei einer längeren Mandatsdauer wahrscheinlicher, eine GCO zu erhalten. Als Begründung wird von den Autoren angeführt, dass Abschlussprüfer in den ersten Jahren ihrer Beschäftigung stärker von den neu gewonnen Mandanten beeinflusst werden. Darüber hinaus führen sie einen Erkenntnisgewinn an, der durch ein besseres Verständnis über die Geschäftsprozesse mit der Zeit erlangt wird. Daneben wird auch von Chi, Myers, Omer und Xie (2017), die taiwanesische Unternehmen im Zeitraum von 1990 bis 2001 analysieren, eine positive Verbindung zwischen der Beschäftigungsdauer und der Wahrscheinlichkeit einer GCO ermittelt.[153] Die Autoren vermuten, dass die Mandatsdauer die Prüfungsqualität verbessert.

Während die vorgenannten Studienergebnisse eine Korrelation zwischen der Beschäftigungsdauer und der Wahrscheinlichkeit der Ausstellung einer GCO aufzeigen, ist bei anderen Studien keine Abhängigkeit erkennbar. Knechel und

[146] Vgl. hierfür und nachfolgend Berglund et al. (2018, S. 5).
[147] Vgl. hierfür und nachfolgend Mo et al. (2015, S. 2-3).
[148] Vgl. hierfür und nachfolgend Read und Yezegel (2016, S. 163-164).
[149] Vgl. Brody und Moscove (1998, S. 35); Copley und Doucet (1993, S. 33); Geiger und Raghunandan (2002, S. 69).
[150] Vgl. Carey und Simnett (2006, S. 673).
[151] Vgl. hierfür und nachfolgend Ye et al. (2011, S. 144-145).
[152] Vgl. hierfür und nachfolgend Geiger und Raghunandan (2002, S. 75).
[153] Vgl. hierfür und nachfolgend Chi et al. (2017, S. 387-388).

Vanstraelen (2007) können bei der Untersuchung von 618 finanziell angespannten Unternehmen in Belgien im Zeitraum von 1992 bis 1996, von denen die eine Hälfte der Unternehmen insolvent ist und die andere Hälfte sich nicht in Insolvenz befindet, keinen Effekt der Beschäftigungsdauer auf die Entscheidung über eine GCO feststellen.[154] Dies bezieht sich allerdings nur auf die Stichprobe der insolventen Unternehmen, wohingegen bei den nicht insolventen Unternehmen geringe Hinweise für eine negative Verbindung vorliegen, d. h. die Typ I Fehlerraten sind bei einer längeren Mandatsdauer niedriger. Weiterhin betrachtet Ratzinger-Sakel (2013) in der Periode von 2005 bis 2009 in Deutschland insgesamt 60 Unternehmen mit einer erstmaligen GCO sowie eine Kontrollgruppe mit 648 bzw. strenge Kontrollgruppe mit 107 finanziell gestressten Unternehmen.[155] Es kann bei Abschlussprüfungen mit einer langen Beschäftigungsdauer (mehr als drei Jahre) nur eine insignifikante Beziehung zwischen dem Niveau der Prüfungshonorare und der Wahrscheinlichkeit der Erteilung einer GCO nachgewiesen werden. Daher liegen keine Belege vor, dass die Beschäftigungsdauer des Abschlussprüfers einen Einfluss auf seine Unabhängigkeit und somit auf seine Entscheidung über die Erteilung einer GCO hat.

3.3.2. Verzögerung der Berichterstattung

Die Hypothese einer Modifizierung des Bestätigungsvermerks in Verbindung mit einer verzögerten Berichterstattung des Abschlussprüfers wird häufig in wissenschaftlichen Untersuchungen aufgegriffen.[156] Unter dem Begriff des Audit Report Lags wird die Anzahl der Tage zwischen dem Datum des Geschäftsjahresendes und dem Tag der Unterzeichnung des Prüfungsberichts durch den Abschlussprüfer bzw. seltener dem Datum der Bekanntgabe der Geschäftsergebnisse verstanden.[157]

McKeown et al. (1991) stellen die Hypothese auf, dass mit einer größeren zeitlichen Verzögerung der Abschlussprüfung die Wahrscheinlichkeit, eine GCO zu erhalten, größer ist.[158] Es wird argumentiert, dass der Zusammenhang aus einem erhöhten zeitlichen Aufwand bei der Prüfung von Unternehmen in einer finanziellen Schieflage bzw. aus der intensiveren Auseinandersetzung mit dem Management bei einer GCO resultiert. Bei der Untersuchung von 134 Unternehmen mit einer Zahlungsunfähigkeit in den Jahren von 1974 bis 1985 sowie einer Kontrollgruppe von 160 Unternehmen wird die Hypothese bestätigt und ein signifikant positiver Zusammenhang zwischen der Länge des Audit Report Lags und der Wahrscheinlichkeit der Ausstellung einer GCO nachgewiesen. Unterstützt werden diese Ergebnisse von Li (2009), die eine positive Verbindung zwischen dem Audit Report Lag und einer GCO feststellt.[159] Für die Analyse werden 1.681 Unternehmen aus dem Jahr 2001 betrachtet, wovon 126 eine erstmalige GCO erhalten haben, sowie 1.780 Unternehmen aus dem Jahr 2003 herangezogen, unter denen sich 108 Unternehmen mit einer erstmaligen GCO befinden. Die Ergebnisse zeigen ebenfalls für beide untersuchten Jahre einen längeren Audit Report Lag für Unternehmen mit einer GCO auf.

Darüber hinaus greift Habib (2013) in seiner Meta-Analyse ebenfalls den Audit Report Lag auf und erwartet eine positive Verbindung zwischen der Länge des Audit Report Lags und der Wahrscheinlichkeit der Erteilung einer GCO.[160] Er führt dabei drei wesentliche Punkte an, die einem längeren Audit Report Lag zugrunde liegen. Einerseits resultiert der größere Zeitraum aus den Verhandlungen zwischen dem Mandanten und dem Abschlussprüfer, in denen versucht wird, Meinungsverschiedenheiten zu lösen. Andererseits muss der Abschlussprüfer nach der Identifikation von Problemen weitere Prüfungshandlungen durchführen und benötigt daher generell mehr Zeit. Des Weiteren kann nach Habib (2013) ein längerer Audit Report Lag auf ein höheres inhärentes Risiko bzw. Kontrollrisiko deuten, aus dem mehr Arbeit für den Abschlussprüfer resultiert. Insgesamt befassen sich 25 Studien mit dem Audit Report Lag und liefern Belege für einen hoch signifikanten und positiven Zusammenhang zwischen dem Audit Report Lag und der Wahrscheinlichkeit der Ausstellung einer GCO.[161] 92 % der Studien können eine positive Verbindung der betrachteten Variablen ermitteln.

Anders als die vorherigen Studien stellen Jiang et al. (2010) keinen signifikanten Zusammenhang fest.[162] Die Stichprobe umfasst die Geschäftsjahre 2004 und 2005 von 1.234 untersuchten Unternehmen, von denen 44 eine GCO erhalten haben. Die Analyse zeigt einen positiven Zusammenhang zwischen der Audit Report Lag Variable und der Wahrscheinlichkeit der Erteilung einer GCO auf, eine signifikante Verbindung kann jedoch nicht entdeckt werden. Die Autoren führen dies auf die hohe Korrelation zwischen der Variablen einer internen Kontrollschwäche und dem Audit Report Lag zurück, da der Abschlussprüfer bei einer wesentlichen Schwäche seine Prüfungstätigkeit erweitert.

Im Gegensatz dazu betrachten Simamora und Hendarjatno (2019) produzierende Unternehmen, die in der Zeit von 2009 bis 2013 an der indonesischen Börse gelistet sind.[163] Eine Verbindung zwischen dem Audit Report Lag und der Wahrscheinlichkeit der Ausstellung einer GCO kann nicht ermittelt werden und daher wird die aufgestellte Hypothese abgelehnt. Die Autoren schließen daraus, dass nicht jedes Unternehmen mit einer GCO einen längeren Prüfungsprozess durchläuft als ein Unternehmen ohne eine GCO. Averio (2021) schließt an diese Studie an und untersucht ebenfalls

[154] Vgl. hierfür und nachfolgend Knechel und Vanstraelen (2007, S. 114 u. 118).
[155] Vgl. hierfür und nachfolgend Ratzinger-Sakel (2013, S. 137-138 u. 150).
[156] Vgl. McKeown et al. (1991, S. 5).
[157] Vgl. DeFond et al. (2002, S. 1256); Goh et al. (2013, S. 977); Li (2009, S.209).
[158] Vgl. hierfür und nachfolgend McKeown et al. (1991, S. 5).
[159] Vgl. hierfür und nachfolgend Li (2009, S. 208 u. 210-211).
[160] Vgl. hierfür und nachfolgend Habib (2013, S. 192).
[161] Vgl. hierfür und nachfolgend Habib (2013, S. 204).
[162] Vgl. hierfür und nachfolgend Jiang et al. (2010, S. 43-44).
[163] Vgl. hierfür und nachfolgend Simamora und Hendarjatno (2019, S. 148 u. 154).

produzierende Unternehmen, die an der indonesischen Börse gelistet sind, jedoch in einem Zeitraum von 2015 bis 2019.[164] Die Studie kommt ebenfalls zu keinem signifikanten Zusammenhang zwischen dem Audit Report Lag und der Wahrscheinlichkeit der Erteilung einer GCO. Averio (2021) sieht den Audit Report Lag nicht als einen geeigneten Indikator an, um abzuschätzen, ob ein Unternehmen in der Fortführung seiner Geschäftstätigkeit gefährdet ist. Es wird argumentiert, dass ein längerer Audit Report Lag sowohl durch Schwierigkeiten im Prüfungsprozess als auch aufgrund von unvollständigen Berichten sowie anderer Faktoren verursacht werden kann.

4. Konsequenzen der Going-Concern-Modifizierung

4.1. Wirkung beim zu prüfenden Unternehmen

4.1.1. Marktreaktion und -bewertung

Der Bestätigungsvermerk des Abschlussprüfers wird mit dem Geschäftsbericht eines Unternehmens veröffentlicht und ist das einzige Dokument, welches der Öffentlichkeit vom Abschlussprüfer nach der Beendigung der Prüfung zur Verfügung steht.[165] Mithilfe der Modifizierung des Bestätigungsvermerks in Form einer GCO kann der Abschlussprüfer sein wahrgenommenes Risiko hinsichtlich der Unternehmensfortführung ausdrücken, wobei die Ausstellung einer GCO zu einer Reaktion bei den Adressaten des Bestätigungsvermerks führt, insbesondere auch in der Bewertung des Unternehmens am Markt.

Bereits frühe Studien in den USA lassen eine negative Reaktion einer GCO auf den Aktienkurs eines Unternehmens erkennen, wobei die negativen abnormalen Erträge (Negative Abnormal Returns) insbesondere in der Periode vor der Bekanntgabe eines qualifizierten Bestätigungsvermerks hoch sind und keine bis schwache Belege für eine geringe negative Reaktion um den Tag der Veröffentlichung selbst gefunden werden.[166] Der geringe Einfluss der Bekanntgabe einer GCO auf den Aktienkurs wird damit begründet, dass die Marktteilnehmer bereits vor der Verkündung über die finanzielle Situation des Unternehmens Bescheid wissen und die Ausstellung einer GCO selbst keine bzw. nur wenig neue Informationen liefert.[167] Die Untersuchung von Blay und Geiger (2001) führt bei der Betrachtung der gesamten Stichprobe von 121 Unternehmen zu keinen signifikant abweichenden abnormalen Erträgen um den Zeitpunkt der Bekanntgabe einer GCO.[168] Allerdings führt die Unterteilung in Unternehmen, die eine GCO erhalten und entweder fortbestehen oder in Konkurs gehen, zu einer negativen Marktreaktion bei den überlebenden Unternehmen, aber zu keiner signifikanten Reaktion bei Unternehmen, die im Anschluss scheitern.

Dahingegen wird von anderen Studien nachgewiesen, dass die Bekanntgabe der GCO einen Informationsgehalt hat und zu einer Marktreaktion führt. Dopuch, Holthausen und Leftwich (1986) können bspw. in ihrer Analyse einen signifikanten negativen Effekt von Mitteilungen in den Wirtschaftsnachrichten über die Qualifizierung eines Bestätigungsvermerks auf den Aktienkurs identifizieren.[169] Ebenso kommt Jones (1996) zu dem Schluss, dass die GCO des Abschlussprüfers den Investoren wichtige Informationen zur Verfügung stellt.[170] Die in seiner Studie betrachtete Stichprobe, die im Vergleich zu anderen früheren Studien eine größere Anzahl von Unternehmen mit einer GCO umfasst, führt zu negativen abnormalen Renditen um den Zeitpunkt der Veröffentlichung des Bestätigungsvermerks vom Abschlussprüfer. Die Untersuchung von Menon und Williams (2010) deckt ebenfalls negative Überschussrenditen (Excess Returns) bei Unternehmen mit einer erstmaligen GCO im Zeitraum von 1995 bis 2006 auf.[171] Darüber hinaus stellen die Autoren fest, dass die negative Reaktion der Investoren bei Unternehmen, die sich mit Finanzierungsschwierigkeiten konfrontiert sehen, stärker ausfällt und eine GCO in diesem Fall den Investoren neue Auskünfte für ihre Entscheidungen bietet.

Daneben untersuchen Menon und Williams (2010) in ihrer Studie den Einfluss der Zusammensetzung der Eigentümerschaft auf die Marktreaktion einer GCO.[172] Bei einem sehr geringen Niveau von institutionellen Investoren wird keine Reaktion auf die Veröffentlichung einer GCO festgestellt, wohingegen bei einem größeren Anteil von institutionellen Beteiligungen die Marktreaktion stärker negativ ist. Die Unterschiede werden von den Autoren auf ein besseres Verständnis der institutionellen Anleger über die Bedingungen der Schuldverschreibungen des Unternehmens zurückgeführt.[173] Bei der Gegenüberstellung von US-Unternehmen mit bzw. ohne einer GCO ermitteln Kaplan, Mowchan und Weisbrod (2014) für einen Zeitraum von fünf Tagen um den Zeitpunkt der Bekanntgabe einer GCO negative abnormale Renditen.[174] Die Autoren merken an, dass die Reaktion auf die GCO abhängig von den institutionellen Investoren ist, die als aktive Marktteilnehmer beschrieben werden und über mehr Ressourcen sowie Informationen verfügen als nichtprofessionelle Anleger. Die negativen abnormalen Erträge sind größer, wenn der Nettoverkauf durch institutionelle Investoren (sog. Institutional Flight) während des Jahres größer ist.

Einige Studien betrachten die Marktreaktion nicht nur in einem Zeitraum direkt nach der Ausstellung einer GCO, sondern auch in einer längeren Periode, wie bspw. die Untersuchung von Taffler, Lu und Kausar (2004) in dem UK, die die Reaktion auf den Aktienkurs über eine Zeit von 12

[164] Vgl. hierfür und nachfolgend Averio (2021, S. 158 u. 161).
[165] Vgl. hierfür und nachfolgend Blay, Geiger und North (2011, S. 77-78).
[166] Vgl. Banks und Kinney (1982, S. 253); Dodd, Dopuch, Holthausen und Leftwich (1984, S. 22-23); Elliott (1982, S. 635-636).
[167] Vgl. Dodd et al. (1984, S. 34); Carson et al. (2013, S. 369).
[168] Vgl. hierfür und nachfolgend Blay und Geiger (2001, S. 223-224).
[169] Vgl. Dopuch et al. (1986, S. 116-117).
[170] Vgl. hierfür und nachfolgend Jones (1996, S. 23-24).
[171] Vgl. hierfür und nachfolgend Menon und Williams (2010, S. 2076).
[172] Vgl. hierfür und nachfolgend Menon und Williams (2010, S. 2104).
[173] Vgl. Menon und Williams (2010, S. 2076).
[174] Vgl. hierfür und nachfolgend Kaplan et al. (2014, 33-34).

Monaten ab dem Monat nach der Informationsveröffentlichung einer GCO analysiert.[175] Von den Autoren wird untersucht, ob Informationen vom Markt direkt vollständig am bzw. um das Datum der Informationsveröffentlichung aufgenommen oder erst mit der Zeit absorbiert werden. Es wird festgestellt, dass der Markt auf die negative Nachricht einer GCO unterreagiert. Informationen werden somit nicht direkt vollständig, sondern vielmehr über eine längere Zeit verarbeitet. Die Unterreaktion wird ebenso von Kausar, Taffler und Tan (2017) in den USA festgestellt, die für Unternehmen mit einer erstmaligen GCO im Zeitraum von 1993 bis 2005 eine signifikant negative Aktienkursrendite von -14 % in den folgenden 12 Monate nach der Bekanntgabe der GCO ermitteln.[176] Die Autoren begründen ihre Befunde u. a. damit, dass hohe Handelskosten keinen Anreiz für Arbitrageure bieten, die Mispricing Anomalie bei einer erstmaligen GCO auszunutzen, weshalb sich die Informationen nicht zeitnah in dem Aktienkurs widerspiegeln. Die Studie von Kausar et al. (2017) wird von Blay, Bryan und Reynolds (2016) repliziert, die für den Zeitraum von 2000 bis 2012 ebenfalls eine signifikante Unterreaktion des Marktes auf die GCO identifizieren.[177] Blay et al. (2016) sind allerdings der Meinung, dass die Berücksichtigung von öffentlich verfügbaren Daten aus den Jahresabschlüssen bei der Zuordnung von Unternehmen mit einer GCO zu einem Unternehmen ohne einer GCO aus der Kontrollgruppe ein passenderer Ansatz wäre, als die Verwendung der Unternehmensgröße und dem Book-to-Market-Verhältnis.[178] Im Kontrast zu den vorherigen Ergebnissen werden bei der Verwendung von Informationen aus Jahresabschlüssen keine signifikanten Kursverluste bei Wertpapieren festgestellt. Die Autoren zeigen darüber hinaus, dass die Abweichungen aus der vorherigen Studie besonders aus kleinen Unternehmen mit einem geringen Niveau an finanziellen Schwierigkeiten und institutionellen Anlegern resultieren.

Der Einfluss einer erstmaligen GCO auf die Gewichtung von Bilanzposten bei der Marktbewertung wird von Blay et al. (2011) betrachtet.[179] Während Subramanyam und Wild (1996) und Barth, Beaver und Landsman (1998) bereits zeigen können, dass der Markt bei der Bewertung von Unternehmen mit finanziellen Schwierigkeiten anstelle der Buchwerte und dem Nettoertrag nur die Buchwerte verwendet, können Blay et al. (2011) diese Befunde ausweiten.[180] Die Ergebnisse ihrer Untersuchung deuten bei der Bewertung auf eine stärkere Gewichtung des Eigenkapital-Buchwerts bei Unternehmen hin, die zum ersten Mal eine GCO ausgestellt bekommen, im Vergleich zu vorherigen Jahren bzw. bei vergleichbaren Unternehmen, die sich in einer finanziellen Notlage befinden und keine GCO erhalten haben.[181] Der Markt schätzt, in Einklang mit einem erhöhten Risiko der Geschäftsaufgabe, die Relevanz der Nettoliquidität, der Forderungen, der langfristigen Vermögenswerte sowie der langfristigen Verbindlichkeiten höher und die Bewertung des Inventars geringer ein. Die Autoren schließen daraus, dass die GCO des Abschlussprüfers für die Marktteilnehmer relevante Informationen bereitstellt.

Den informativen Wert einer GCO können ebenfalls Kausar und Lennox (2017) nachweisen, die den Informationsgehalt einer vorangegangenen GCO hinsichtlich der Differenz zwischen dem Buchwert von Vermögensgegenständen und dem zukünftig erzielbaren Wert bei einer Liquidation aufzeigen.[182] Die GCO kann somit ein Warnsignal für die Adressaten des Jahresabschlusses darstellen, dass die Liquidationswerte bei einer Insolvenz wesentlich kleiner als die Buchwerte sein können.

4.1.2. Kapitalkosten

Mit der Ausstellung einer GCO durch den Abschlussprüfer werden erhebliche Bedenken einer möglichen Gefährdung der Unternehmensfortführung geäußert, was von Eigen- und Fremdkapitalgebern als negatives Signal gedeutet werden kann.[183] Daher wird in der Prüfungsforschung u. a. argumentiert, dass ein qualifizierter Bestätigungsvermerk mit dem systematischen Risiko eines Unternehmens verbunden und das Schuldenrisiko erhöht sein kann sowie die Ausstellung einer GCO zu höheren Kapitalkosten für das Unternehmen führt.[184]

Die negativen Folgen einer GCO auf die Eigenkapitalkosten werden bspw. von Amin et al. (2014) untersucht, die 106 Unternehmen mit einer GCO und 106 Unternehmen in finanzieller Schieflage ohne einer GCO in den Jahren von 2000 bis 2010 betrachten.[185] Es kann dabei ein signifikant positiver Zusammenhang zwischen einer GCO und den Eigenkapitalkosten gefunden werden.[186] Die Eigenkapitalkosten erhöhen sich durch die Erteilung einer GCO abhängig von der angewandten Modellspezifikation um 3,3 % bis 5,7 %. Eine GCO informiert die Eigenkapitalgeber somit über die finanzielle Lage und eine mögliche Insolvenz des Unternehmens, worauf die Investoren aufgrund des erhöhten Risikos eine erhöhte Rendite fordern und das Unternehmen sich dadurch mit höheren Eigenkapitalkosten konfrontiert sieht.

Darüber hinaus zeigen Chen, He, Ma und Stice (2016) die Auswirkungen einer GCO auf die Kreditverträge eines Unternehmens auf.[187] In ihrer Analyse von Unternehmen im Zeitraum von 1992 bis 2009 stellen sie fest, dass sich der Kreditaufschlag (Loan Spread) nach einer GCO um 107 Basispunkte erhöht und die betroffenen Unternehmen nicht nur einen höheren Zinssatz für das gestiegene Risiko zahlen müssen,

[175] Vgl. hierfür und nachfolgend Taffler et al. (2004, S. 264-265).
[176] Vgl. hierfür und nachfolgend Kausar et al. (2017, S.213-214).
[177] Vgl. Blay et al. (2016, S. 15).
[178] Vgl. hierfür und nachfolgend Blay et al. (2016, S. 21-22).
[179] Vgl. Blay et al. (2011, S. 98-100).
[180] Vgl. Barth et al. (1998, S. 31); Blay et al. (2011, S. 98); Subramanyam und Wild (1996, S. 269).
[181] Vgl. hierfür und nachfolgend Blay et al. (2011, S. 100).
[182] Vgl. hierfür und nachfolgend Kausar und Lennox (2017, S. 899).
[183] Vgl. Amin, Krishnan und Yang (2014, S. 5).
[184] Vgl. Amin et al. (2014, S. 2); Fargher und Wilkins (1998, S. 845); J. Francis, LaFond, Olsson und Schipper (2004, S. 969-970).
[185] Vgl. Amin et al. (2014, S. 9).
[186] Vgl. hierfür und nachfolgend Amin et al. (2014, S. 32).
[187] Vgl. hierfür und nachfolgend Chen et al. (2016, S. 4-5).

sondern die Kreditgeber zudem das Kreditvolumen reduzieren und eine signifikante Kürzung der Laufzeit von Krediten, die nach einer GCO erteilt werden, erkennbar ist. Ebenfalls zeigt die Studie einen Rückgang der finanziellen Vereinbarungen in Kreditverträgen um 9,1 % und einen Anstieg der allgemeinen Vereinbarungen um 12,3 %, was darauf hindeutet, dass die Kreditgeber sich weniger auf die finanziellen Vereinbarungen verlassen wollen, nachdem die Qualität der Finanzberichterstattung durch die GCO in Frage gestellt wurde. Chen et al. (2016) kommen zu dem Schluss, dass den Kreditgebern wichtige Informationen zum Kreditrisiko des Unternehmens zur Verfügung gestellt und in der Kreditvergabe berücksichtigt werden.

Dahingegen schlussfolgern Niemi und Sundgren (2012), dass Kreditgeber nicht ausschließlich nur öffentlich verfügbare Informationen in die Kreditvergabeentscheidung mit einbeziehen, sondern auch andere Informationsquellen nutzen.[188] Die Studie umfasst mehr als fünfzigtausend Beobachtungen von kleinen und mittleren Unternehmen in Finnland, wobei rund zweitausend Unternehmen einen modifizierten Bestätigungsvermerk aufweisen, der eine Red Flag Opinion, vergleichbar mit einer GCO, beinhaltet. Es wird untersucht, ob die Verfügbarkeit der Kredite von institutionellen Kreditgebern durch eine Modifizierung des Bestätigungsvermerks verringert und stattdessen auf Handelskredite zurückgegriffen wird. Für die Red Flag Opinion können allerdings keine Belege für einen Zusammenhang von einer gestiegenen Verwendung von Handelskrediten im Verhältnis zu institutionellen Krediten gefunden werden.

Die Akteure an den Kapitalmärkten nutzen als zusätzliche Informationsquelle für ihre Entscheidungen die Auskünfte von Ratingagenturen, die die Kreditwürdigkeit von Unternehmen bewerten.[189] Eine der wenigen Untersuchungen, die den Zusammenhang zwischen einer GCO und dem Kreditrating durch eine Ratingagentur betrachtet, ist die Studie von Feldmann und Read (2013). In die Analyse werden 152 Unternehmen, die im Zeitraum von 2000 bis 2009 Insolvenz angemeldet haben und bei denen die Kreditwürdigkeit durch die Ratingagentur Standard & Poor‘s (S&P) bewertet wurde, einbezogen.[190] Die Autoren ermitteln bei 56 % der Unternehmen, die eine GCO erhalten haben, eine Herabsetzung der Bonitätsbewertung in dem auf die GCO folgenden Monat, während nur bei 21 % der Unternehmen mit einem uneingeschränkten Bestätigungsvermerk eine Abwertung erfolgt. Die Studie von Feldmann und Read (2013) wird durch die Untersuchung von Strickett, Hay und Lau (2021) erweitert, die neben S&P die Ratingagentur Moody’s aufnimmt und insolvente Unternehmen in der Zeit von 2002 bis 2013 analysiert.[191] Die Stichprobe von S&P mit einem Umfang von 149 Unternehmen zeigt bei 68 % der Unternehmen mit einer GCO eine Abwärtsbewegung in dem der GCO anschließenden Monat und nur eine Herabsetzung der Kreditwürdigkeit bei 39 % der Unternehmen ohne eine GCO. Dahingegen ist der Unterschied bei Moody’s, wo die Stichprobe 147 Unternehmen enthält, insignifikant. Während bei 24 % der Unternehmen mit einer GCO in dem auf die GCO folgenden Monat eine Abwertung erfolgt, liegt dieser Wert bei Unternehmen mit einem nicht modifizierten Bestätigungsvermerk bei 21 %. Die geringe Reaktion von Moody’s auf die GCO wird von den Autoren auf eine andere Eigentümerstruktur der Ratingagentur zurückgeführt.[192]

4.1.3. Selbsterfüllende Prophezeiung

Mit der Berichterstattung des Abschlussprüfers über die Fortführung der Unternehmenstätigkeit wird stets die Möglichkeit einer selbsterfüllenden Prophezeiung (Self-Fulfilling Prophecy) diskutiert.[193] Die selbsterfüllende Prophezeiung stellt das Eintreten eines vorhergesagten Ereignisses aufgrund der vorherigen Veröffentlichung einer Prognose dar. Der Argumentation der Literatur folgend, kommt es durch die Bekanntgabe einer GCO, die von den Stakeholdern eines Unternehmens als negatives Signal bzw. Insolvenzprognose des Abschlussprüfers angesehen wird, zu einem Vertrauensverlust in der Öffentlichkeit. Als Konsequenz wird hierdurch die negative Unternehmensentwicklung durch bspw. die Kündigung von Kreditverträgen oder das Nichtzustandekommen von Aufträgen gefördert. Der Abschlussprüfer wäre demnach durch die Erteilung einer GCO für einen Unternehmenszusammenbruch mitverantwortlich. Daher kann die Entscheidung des Abschlussprüfers über die Ausstellung einer GCO durch die selbsterfüllende Prophezeiung beeinflusst werden.[194] Heinrichs (2018) hebt hervor, dass beim Vorhandensein einer selbsterfüllenden Prophezeiung eine GCO, die sich ex ante als fehlerhaft herausstellt, bei einem gesunden Unternehmen der Grund für die Einstellung der Geschäftstätigkeit darstellen kann.[195]

Wissenschaftliche Untersuchungen kommen in der Frage, ob eine selbsterfüllende Prophezeiung bei einer GCO vorliegt oder dieses Phänomen nicht auftritt, zu verschiedenen Ergebnissen. Louwers, Messina und Richard (1999) untersuchen US-Unternehmen mit einer erstmaligen GCO im Zeitraum von 1984 bis 1991 auf die Wahrscheinlichkeit eines Unternehmenszusammenbruchs in den folgenden zehn Jahren.[196] Nach dem Erhalt einer GCO überleben Unternehmen im Durchschnitt vier Jahre und nur 27 % der betrachteten Unternehmen scheitern in dem auf die GCO folgenden Jahr. Dennoch ist die Insolvenzwahrscheinlichkeit insbesondere in den ersten beiden Jahren hoch, nimmt aber anschließend ab. Die Ergebnisse deuten darauf hin, dass die selbsterfüllende Prophezeiung höchstens in den ersten Jahren existiert und in den folgenden Jahren, wenn überhaupt, nur einen sehr

[188] Vgl. hierfür und nachfolgend Niemi und Sundgren (2012, S. 768-769).
[189] Vgl. hierfür und nachfolgend Feldmann und Read (2013, S. 346).
[190] Vgl. hierfür und nachfolgend Feldmann und Read (2013, S. 358).
[191] Vgl. hierfür und nachfolgend Strickett et al. (2021, S. 11).
[192] Vgl. Strickett et al. (2021, S. 13).
[193] Vgl. hierfür und nachfolgend Adam (2007, S. 62); Marten et al. (2020, S. 613-614).
[194] Für einen Überblick über Studien, die den Einfluss einer möglichen selbsterfüllenden Prophezeiung auf die GCO-Entscheidung des Abschlussprüfers untersuchen, siehe Heinrichs (2018, S. 166).
[195] Vgl. Heinrichs (2018, S. 2).
[196] Vgl. hierfür und nachfolgend Louwers et al. (1999, S. 811 u. 814-815).

geringen Effekt aufweist. Einhergehend mit diesem Befund kann Nogler (2004) für die Periode von 1985 bis 1991 in den USA keine Anzeichen für ein mit der GCO verbundenes Langzeit-Stigma finden.[197] Er analysiert Unternehmen, die eine GCO erhalten haben und im folgenden Jahr wieder einen uneingeschränkten Bestätigungsvermerk verzeichnen. Dabei stellt er fest, dass über 70 % dieser Unternehmen in einem Zeitraum von 12 bis 18 Jahren den Aktionären einen Mehrwert schaffen und somit eine GCO, die im folgenden Jahr nicht mehr ausgestellt wurde, auf lange Sicht keinen Einfluss auf die Unternehmenstätigkeit hat. Die Studie von Gerakos, Hahn, Kovrijnykh und Zhou (2016), die GCOs und Insolvenzanmeldungen von Unternehmen, die im vorherigen Jahr keine GCO erhalten haben, auswertet, stellt lediglich eine um 0,84 % höhere Wahrscheinlichkeit für eine Insolvenz in Folge einer GCO fest.[198] Der geringe Anstieg der Wahrscheinlichkeit einer Insolvenz deutet somit nicht auf einen wesentlichen Einfluss der GCO auf eine Insolvenz hin.

In dem UK vergleichen Citron und Taffler (2001) Unternehmen, die zwischen 1986 und 1993 eine GCO erhalten haben, mit Unternehmen, denen keine GCO ausgestellt wurde.[199] Zwar ermitteln die Autoren bis zur Veröffentlichung des nächsten Jahresabschlusses einen signifikanten Unterschied hinsichtlich der Insolvenzrate zwischen den Unternehmen mit einem qualifizierten (17 %) und einem nichtqualifizierten (8 %) Bestätigungsvermerk, womit eine selbsterfüllende Prophezeiung belegt werden könnte. Allerdings wird kein signifikanter Unterschied gefunden, wenn das Modell um eine Variable erweitert wird, die die finanziellen Schwierigkeiten eines Unternehmens abbildet. Eine selbsterfüllende Prophezeiung kann dadurch nicht mehr nachgewiesen werden. Vielmehr deuten Citron und Taffler (2001) an, dass doppelt so viele Unternehmen größere Restrukturierungen vornehmen als zahlungsunfähig werden.[200] Eine GCO hat damit der Hypothese der Self-Defeating Prophecy entsprechend, die eine Falsifizierung bzw. Selbstzerstörung der Prognose durch ein bewusstes Handeln des Managements und der Stakeholder gegen die Bestandsgefährdung annimmt, einen positiven Effekt auf das Überleben eines Unternehmens. Bei Erhalt einer GCO können somit rechtzeitig Maßnahmen zur Rettung des Unternehmens ergriffen werden.

Ebenfalls führt die Studie von Carey et al. (2008) zu einer Ablehnung der Existenz einer selbsterfüllenden Prophezeiung.[201] In ihrer Analyse australischer Unternehmen in der Periode von 1994 bis 1997 wurde festgestellt, dass 11,8 % der Unternehmen mit einer GCO und 10,3 % der Unternehmen ohne einer GCO innerhalb der gleichen Zeitspanne von zwei Jahren gescheitert sind. Dieser Unterschied ist nicht signifikant und stellt daher keinen Beleg für eine selbsterfüllende Prophezeiung, zumindest in Australien, dar.

Dahingegen finden Gaeremynck und Willekens (2003) in Belgien bei der Untersuchung von 228 Unternehmen in den Jahren von 1995 und 1996 deutliche Hinweise für eine selbsterfüllende Prophezeiung.[202] Es kann nachgewiesen werden, dass ein positiver Effekt zwischen der Erteilung einer GCO und der Insolvenzwahrscheinlichkeit besteht. Demgegenüber zeigt jedoch die Teilstichprobe von Unternehmen, die sich freiwillig liquidieren, keine signifikanten Ergebnisse und lässt somit keine Anzeichen für eine selbsterfüllende Prophezeiung erkennen. Auch die belgische Studie von Vanstraelen (2003), die Unternehmen in dem Zeitraum von 1992 bis 1996 betrachtet, kommt zu dem Schluss, dass eine GCO die Wahrscheinlichkeit einer Insolvenz signifikant erhöht.[203] Die Ergebnisse deuten auf die Existenz einer selbsterfüllenden Prophezeiung hin und sind sowohl bei einer erstmaligen GCO als auch bei einer wiederholten Ausstellung der GCO zu finden.

Für den deutschen Prüfungsmarkt analysiert Frey (2014) den Einfluss der GCO auf die Wahrscheinlichkeit einer Unternehmensbeendigung bzw. -insolvenz in den Jahren von 2006 bis 2010.[204] Sowohl bei der Verwendung der Unternehmensbeendigung als auch der Unternehmensinsolvenz als abhängige Variable ist ein signifikant positiver Zusammenhang zwischen der Ausstellung einer GCO und der Wahrscheinlichkeit einer Unternehmensbeendigung bzw. -insolvenz erkennbar, sodass die Existenz einer selbsterfüllenden Prophezeiung angenommen werden kann. Allerdings wird dieser Zusammenhang bei der Unternehmensbeendigung in der ersten Untersuchungsauswahl nicht nachgewiesen. Hier wird argumentiert, dass in dieser Stichprobe die Unterschiede bezüglich der VFE-Lage der Unternehmen zwischen der Untersuchungs- und Kontrollgruppe größer sind und das verwendete Matching-Verfahren bei der betrachteten Forschungsfrage kein angemessenes Instrument darstellt.

4.2. Effekt beim Abschlussprüfer

4.2.1. Abschlussprüferwechsel

Die Erteilung einer GCO kann als Konsequenz für den Abschlussprüfer die Beendigung des Mandats durch das zu prüfende Unternehmen zur Folge haben, womit das Unternehmen beabsichtigt, eine erneute GCO zu vermeiden.[205] Unter dem Begriff des Opinion Shoppings ist der Versuch zu verstehen, einen uneingeschränkten Bestätigungsvermerk zu erhalten. Der Mandant droht dem aktuellen Abschlussprüfer einen Wechsel an bzw. lotet bei anderen Abschlussprüfern aus, wie diese in relevanten Sachverhalten reagieren würden und wechselt dann zu dem Abschlussprüfer, dessen Einstellung sich am meisten mit der des Unternehmens überschneidet.

[197] Vgl. hierfür und nachfolgend Nogler (2004, S. 681).
[198] Vgl hierfür und nachfolgend Gerakos et al. (2016, S. 1 u. 14-15).
[199] Vgl. hierfür und nachfolgend Citron und Taffler (2001, S. 357-358).
[200] Vgl. hierfür und nachfolgend Citron und Taffler (2001, S. 360); Ruhnke (2003, S. 263).
[201] Vgl. hierfür und nachfolgend Carey et al. (2008, S. 74-75).
[202] Vgl. hierfür und nachfolgend Gaeremynck und Willekens (2003, S. 75-76).
[203] Vgl. hierfür und nachfolgend Vanstraelen (2003, S. 243-246).
[204] Vgl. hierfür und nachfolgend Frey (2014, S. 188-189).
[205] Vgl. hierfür und nachfolgend Chow und Rice (1982, S. 327); Marten et al. (2020, S. 202).

Das Ereignis des Abschlussprüferwechsels wird von Chow und Rice (1982) untersucht, die hierfür Unternehmen mit einem qualifizierten Bestätigungsvermerk im Jahr 1973 sowie einem Wechsel des Abschlussprüfers zwischen den Jahren 1973 und 1974 betrachten (der Grund für den qualifizierten Bestätigungsvermerk wird nicht weiter differenziert).[206] Die Ergebnisse zeigen, dass Unternehmen nach einem qualifizierten Bestätigungsvermerk den Abschlussprüfer häufiger wechseln. Es werden allerdings keine Belege für einen systematischen Wechsel von einem Abschlussprüfer mit einer hohen Tendenz zur Erteilung von qualifizierten Bestätigungsvermerken zu einem Abschlussprüfer mit einer niedrigen Tendenz entdeckt.[207] Darüber hinaus kann keine signifikant geringere Anzahl an nicht-qualifizierten Bestätigungsvermerken gefunden werden, wenn der Abschlussprüfer gewechselt wird.

Dahingegen beleuchten J. Krishnan und Stephens (1995) in ihrer Studie unterschiedliche Formen des Bestätigungsvermerks, wobei insbesondere auch die GCO berücksichtigt wird, die im Zeitraum von 1986 bis 1987 bzw. von 1987 bis 1988 einen Anteil von 5,47 % bzw. 8,7 % der betrachteten Unternehmen ausmacht.[208] Die Autoren stellen fest, dass die Rate der Unternehmen, die nach einer GCO den Abschlussprüfer wechseln, mit 22,61 % deutlich höher ist als bei einem uneingeschränkten Prüfungsurteil mit nur 5,02 %.[209] Ein Anhaltspunkt für ein erfolgreiches Opinion Shopping kann jedoch nicht identifiziert werden, da das Unternehmen mit einem Abschlussprüferwechsel sowohl vom vorherigen als auch dem nachfolgenden Abschlussprüfer eine ähnliche Behandlung erhält.[210] Es wird argumentiert, dass der Wechsel des Abschlussprüfers nicht durch ein Opinion Shopping induziert ist, sondern vielmehr andere Gründe den Auslöser für den Abschlussprüferwechsel darstellen. Des Weiteren führen die Autoren an, dass, wenn die Wechselraten durch das Opinion Shopping begründet wären, die eingeführten Maßnahmen gegen ein Opinion Shopping Wirkung zeigen und die Behandlung der Unternehmen daher ähnlich bleibt.

Die Auswirkung einer GCO auf den Abschlussprüferwechsel wird ebenfalls in der bereits genannten Studie von Carey et al. (2008) anhand von 68 australischen Unternehmen mit einer erstmaligen GCO betrachtet.[211] Zum Vergleich wird eine Kontrollgruppe mit Unternehmen gebildet, die keine GCO ausgestellt bekommen haben, sich aber in einer finanzieller Schieflage befinden, und anschließend anhand von Kennzahlen mit der Stichprobe der GCO-Unternehmen gematcht. Während 16,7 % der Unternehmen mit einer GCO den Abschlussprüfer innerhalb eines Jahres wechseln, sind es bei Unternehmen ohne eine GCO nur 9,8 %.[212] Dieser Unterschied stellt sich allerdings als nicht signifikant heraus. Wird hingegen ein Logit-Modell verwendet, ist die Variable der GCO signifikant positiv, womit eine höhere Tendenz eines Abschlussprüferwechsels bei Unternehmen mit einer erstmaligen GCO im Vergleich zu Unternehmen ohne eine GCO nachgewiesen werden kann. Australische Abschlussprüfer sehen sich somit nach einer erstmaligen GCO mit einer höheren Wahrscheinlichkeit des Verlustes des Mandanten konfrontiert.

Carcello und Neal (2003) bringen den Wechsel des Abschlussprüfers nach einer GCO mit der Zusammenstellung des Audit Committees in Verbindung und analysieren Unternehmen, die zwischen 1988 und 1999 erstmalig eine GCO erhalten und ihren Abschlussprüfer vor dem nächsten Jahresabschluss gewechselt haben.[213] Die Untersuchungen zeigen, dass, je größer der Anteil der Affiliated Direktoren im Audit Committee ist, die Wahrscheinlichkeit eines Abschlussprüferwechsels nach einer GCO steigt.[214] Ebenso erhöht sich die Wahrscheinlichkeit für einen Wechsel nach einer GCO, wenn die Mitglieder des Audit Committees mehr Anteile am Unternehmen besitzen. Außerdem deuten die Ergebnisse darauf hin, dass ein Opinion Shopping bei Unternehmen mit einem größeren Anteil an Affiliated Direktoren mit einer höheren Wahrscheinlichkeit erfolgreich ist, da 37,1 % der Unternehmen mit einem Anteil an Affiliated Direktoren oberhalb des Medians nach einem Wechsel ein uneingeschränktes Prüfungsurteil erhalten, während es bei den Unternehmen mit einem Anteil unterhalb des Medians nur 18,5 % sind. Hieran anknüpfend bindet Kim (2021) das Management in die Untersuchung ein und stellt einen signifikant positiven Zusammenhang zwischen einem übermäßigen Selbstvertrauen des Managements und dem Abschlussprüferwechsel nach einer GCO fest.[215] Für die Messung des übermäßigen Selbstvertrauens wird die optionsbasierte Methode als Proxy verwendet. Demnach wird ein Manager als übermäßig selbstbewusst angesehen, wenn die Optionsausübung verzögert wurde, obwohl die Option zu dem Zeitpunkt In-The-Money war, also einen Gewinn bei Ausübung erzielt hätte. Die Ergebnisse legen dar, dass die Verbindung zwischen dem übermäßigen Selbstvertrauen des Managements nach einer GCO und der Wahrscheinlichkeit eines Abschlussprüferwechsels dann stärker ist, wenn das Management im Vergleich zum Audit Committee mächtiger ist. Diese Erkenntnisse deuten darauf hin, dass die Besetzung bzw. Eigenschaften des Managements und des Audit Committees einen Einfluss auf den Wechsel des Abschlussprüfers haben.

4.2.2. Finanzielle Folgen und rechtliche Auswirkungen

Durch den Abschlussprüferwechsel nach einer GCO verliert der Abschlussprüfer nicht nur seinen Prüfungsauftrag bei dem Unternehmen, sondern kann mit dem Mandanten auch keinen Umsatz mehr generieren.[216] Die bereits erwähnte Studie von Carey et al. (2008) zeigt, dass der Anteil der

[206] Vgl. hierfür und nachfolgend Chow und Rice (1982, S. 328).
[207] Vgl. hierfür und nachfolgend Chow und Rice (1982, S. 332-333).
[208] Vgl. J. Krishnan und Stephens (1995, S. 187).
[209] Vgl. J. Krishnan und Stephens (1995, S. 189).
[210] Vgl. hierfür und nachfolgend J. Krishnan und Stephens (1995, S. 196-197).
[211] Vgl. hierfür und nachfolgend Carey et al. (2008, S. 66-67).
[212] Vgl. hierfür und nachfolgend Carey et al. (2008, S. 72-73).
[213] Vgl. Carcello und Neal (2003, S. 101).
[214] Vgl. hierfür und nachfolgend Carcello und Neal (2003, S. 113).
[215] Vgl. hierfür und nachfolgend Kim (2021, S. 5 u. 10-11).
[216] Vgl. hierfür und nachfolgend Carey et al. (2008, S. 73-74).

verlorenen Prüfungshonorare beim Wechsel eines Unternehmens mit einer GCO signifikant größer ist als bei einem Wechsel von Unternehmen ohne einer GCO.

Berglund (2020) erläutert zunächst, dass das Management für eine hohe Prüfungsqualität einen Aufschlag zahlen würde, dagegen Abschlussprüfer, bei denen Typ II Fehler aufgetreten sind, ihre Prüfungshonorare senken müssten, damit sie ihre Mandanten halten können.[217] Demgegenüber argumentiert er, dass solche Abschlussprüfer einen Honoraraufschlag verlangen können, wenn ein Unternehmen in der Entscheidung über die Wahl des Abschlussprüfers durch ein Opinion Shopping beeinflusst ist. In seiner Analyse kann Berglund (2020) allerdings keine Belege für einen signifikanten Zusammenhang zwischen Abschlussprüfern mit Typ II Fehlern und den Prüfungshonoraren finden.

Die experimentelle Studie von Christensen, Glover, Omer und Shelley (2016), deren Umfrage im Jahr 2012 von 93 Wirtschaftsprüfungsexperten (Partner und Senior Manager) sowie 102 Investoren beantwortet wurde, beschäftigt sich dagegen mit dem Typ I Fehler.[218] Die Untersuchung zeigt, dass bei einer Fehleinschätzung, bei der ein Mandant eine GCO erhält, aber in der folgenden Periode nicht Insolvenz anmeldet, dies sowohl von den Wirtschaftsprüfungsexperten als auch den Investoren als ein Zeichen für eine hohe Prüfungsqualität gesehen wird und der Abschlussprüfer keine Beeinträchtigung seiner Reputation befürchten muss. Die Typ II Fehler werden dahingegen von den Probanden negativ eingeordnet und führen zu einer durch den Abschlussprüfer und die Investoren geringeren wahrgenommen Prüfungsqualität.

Die rechtlichen Auswirkungen einer Modifizierung des Prüfungsurteils zeigen Carcello und Palmrose (1994) in ihrer Studie in den Jahren von 1972 bis 1992, die 655 Unternehmen umfasst, die sich in Insolvenz befinden und von denen sich 118 Unternehmen einer Rechtsstreitigkeit mit dem Abschlussprüfer stellen müssen.[219] Die Ergebnisse deuten darauf hin, dass modifizierte Bestätigungsvermerke tendenziell zu geringeren Ansprüchen nach Rechtsstreitigkeiten führen und die modifizierten Bestätigungsvermerke zur Verringerung der Forderungen geeignet sind. Allerdings schützt die Modifizierung des letzten Bestätigungsvermerks vor einer Insolvenz nicht vor einer rechtlichen Auseinandersetzung.

An das Thema anknüpfend führen Kaplan und Williams (2013) für den Zeitraum von 1986 bis 2009 eine Untersuchung von finanziell angeschlagenen Unternehmen durch, die sich mit einer Wertpapier-Sammelklage konfrontiert sehen.[220] Die Autoren verwenden neben einer Probit-Analyse auch einen simultanen Gleichungsansatz, um die Endogenität besser kontrollieren zu können.[221] Während die Variable für die GCO in der Probit-Analyse nicht signifikant ist, kann diese im simultanen Gleichungsansatz als signifikant und negativ nachgewiesen werden. Die Ergebnisse weisen somit auf eine geringere Wahrscheinlichkeit einer Rechtsstreitigkeit gegen den Abschlussprüfer hin, wenn dieser eine GCO erteilt hat. Darüber hinaus kann festgestellt werden, dass nach der Ausstellung einer GCO das Urteil einer Klage gegen den Abschlussprüfer weniger schwerwiegend ist. Die Wahrscheinlichkeit für die Zahlung eines großen finanziellen Betrags ist nach der Erteilung einer GCO geringer und die Höhe der finanziellen Beilegung fällt für den Abschlussprüfer niedriger aus.

Zu einem konträren Ergebnis kommen Eutsler, Nickell und Robb (2016), die mit der Durchsicht von Auditing and Accounting Enforcement Releases (AAERs), die von der SEC im Zeitraum von 1995 bis 2012 ausgestellt wurden, einen anderen Ansatz wählen.[222] Aus den Unternehmen, die mindestens einen Jahresabschluss mit Betrug aufweisen, wurde die finale Stichprobe mit 314 Unternehmen erstellt, wovon 34 Unternehmen eine GCO erhalten haben sowie 12 Unternehmen sowohl in Strafverfolgungsmaßnahmen gegen den Abschlussprüfer verwickelt waren als auch eine GCO ausgestellt bekommen haben. Die Analyse kommt zu dem Ergebnis, dass die getestete Variable GCO signifikant und positiv ist.[223] Daraus kann gedeutet werden, dass für den Abschlussprüfer die Wahrscheinlichkeit für eine Sanktionierung nach einem Betrug größer ist, wenn der Abschlussprüfer die finanzielle Schieflage eines Unternehmens wahrgenommen und eine GCO erstellt hat. Die GCO wird von Regulierungsbehörden bei der Untersuchung von Betrug in der Finanzberichterstattung anders bewertet als die Erwägung einer Klage von Investoren nach einer Insolvenz.

Das Experiment von Wright und Wright (2014) zeigt darüber hinaus, dass die Aufnahme von Informationen über das Beurteilungsverfahren des Abschlussprüfers hinsichtlich der Entscheidung einer GCO einen Einfluss auf Investoren haben kann.[224] Für das Experiment haben 72 nicht-professionelle Investoren, ersatzweise dargestellt durch MBA Studenten, Informationen zu einem Unternehmen und der Verantwortung des Abschlussprüfers für die Beurteilung der Unternehmensfortführung erhalten.[225] Darunter haben sich drei Hinweise für eine Fortführung der Geschäftstätigkeit sowie drei Informationen, die Zweifel an der Unternehmensfortführung aufrufen, befunden. Das betrachtete Unternehmen hat ein uneingeschränktes Prüfungsurteil erhalten und neun Monate nach dem Bestätigungsvermerk des Abschlussprüfers Insolvenz angemeldet. Von den 72 Teilnehmern haben 37 Probanden Informationen über den Beurteilungsprozess des Abschlussprüfers erhalten und 35 Probanden haben diese Informationen nicht erhalten. Die Ergebnisse zeigen eine bessere Bewertung des Abschlussprüfers, wenn die zusätzlichen Informationen mitgeteilt wurden.[226] Bei vier von fünf Variablen ist ein signifikanter Unterschied zwischen den beiden Gruppen vorhanden, sodass Investoren mit den zusätzlichen Infor-

[217] Vgl. hierfür und nachfolgend Berglund (2020, S. 1927).
[218] Vgl. hierfür und nachfolgend Christensen et al. (2016, S. 1650 u. 1663).
[219] Vgl. hierfür und nachfolgend Carcello und Palmrose (1994, S. 8 u. 27).
[220] Vgl. Kaplan und Williams (2013, S. 206).
[221] Vgl. hierfür und nachfolgend Kaplan und Williams (2013, S. 222-224).
[222] Vgl. hierfür und nachfolgend Eutsler et al. (2016, S. 384).
[223] Vgl. hierfür und nachfolgend Eutsler et al. (2016, S. 387).
[224] Vgl. Wright und Wright (2014, S. 35).
[225] Vgl. hierfür und nachfolgend Wright und Wright (2014, S. 39-41).
[226] Vgl. hierfür und nachfolgend Wright und Wright (2014, S. 36-37).

mationen die Leistung des Abschlussprüfers deutlich weniger negativ bewerten.

5. Kritische Würdigung

Die GCO des Abschlussprüfers im Bestätigungsvermerk stellt ein wichtiges Instrument der Kommunikation zwischen dem Abschlussprüfer und der Öffentlichkeit dar. Die enthaltenen Informationen dienen als Entscheidungsgrundlage für die Adressaten des Jahresabschlusses und geben Auskunft über mögliche wesentliche Unsicherheiten bei der Fortführung der Geschäftstätigkeit eines Unternehmens. Zahlreiche Studien befassen sich mit den Determinanten der GCO und zeigen ebenso die Konsequenzen der Erteilung einer GCO auf. Bei der Auswertung der Studien ist jedoch zu berücksichtigen, dass es hinsichtlich der GCO zu Unterschieden in den gesetzlichen und normativen Vorgaben zwischen den Ländern kommt und mit dem Code Law und Common Law verschiedene Rechtssysteme vorliegen. Daher ist eine direkte Übertragung von Forschungsergebnissen von einem Rechtsraum auf einen anderen u. U. nicht ohne weiteres möglich und muss unter Beachtung gewisser Limitationen erfolgen. Zudem basieren vereinzelte Studienergebnisse auf einem geringen Stichprobenumfang, sodass eine geringe externe Validität bei der Interpretation der Befunde beachtet werden muss.

Die Determinanten der GCO können in verschiedene Einflussfaktoren unterteilt werden, wobei eine Differenzierung insbesondere in die Faktoren vom zu prüfenden Unternehmen und vom Abschlussprüfer sowie deren Beziehung zueinander erfolgen kann. Ergänzend liegt mit den Umweltfaktoren – hierzu zählen bspw. Prüfungsstandards, das regulatorische Umfeld, Finanz- und Wirtschaftskrisen, nationale Unterschiede, die Marktstruktur und der Wettbewerb – ein weiterer Informationskreis vor, der allerdings in dieser Arbeit nicht im Detail erarbeitet werden kann.

Die Faktoren, die vom zu prüfenden Unternehmen ausgehen, befassen sich mit den finanziellen Kennzahlen, die dem Jahresabschluss entnommen werden. Hier zeigt sich, dass die Ausstellung einer GCO für Unternehmen, die weniger profitabel, weniger liquide und kleiner sind sowie ein höheres Level an Leverage, eine geringere Rendite und eine größere Renditevolatilität aufweisen, wahrscheinlicher ist. Die Ergebnisse einzelner Studien, die keinen signifikanten Einfluss dieser Kennzahlen auf die Wahrscheinlichkeit einer GCO nachweisen können, sind auf die unterschiedlichen Untersuchungsansätze und -methoden zurückzuführen. Ungeklärt bleibt die Frage, ob die Kennzahlen als Grundlage für die Entscheidung des Abschlussprüfers dienen oder es sich um Zusammenhänge handelt, die durch andere Ursachen bedingt sind.

Ebenso haben Entscheidungen, die vom Management getroffen werden, einen Einfluss auf die Ausstellung einer GCO. Während Pläne für die Ausgabe von neuen Aktien sowie die Aufnahme zusätzlicher Mittel zu einer geringeren Wahrscheinlichkeit einer GCO führen und damit als ein positives Signal nach außen gewertet werden können, haben die Kürzung von Ausgaben bzw. die Veräußerung von Vermögenswerten, insbesondere bei liquiditätsschwachen Unternehmen, eine negative Wirkung. Strategische Maßnahmen können bei einem sowohl kurz- als auch langfristigen positiven Effekt auf den Cashflow einen mildernden Faktor darstellen und die Wahrscheinlichkeit einer GCO reduzieren, wohingegen nur langfristige Erfolge die Wahrscheinlichkeit einer GCO erhöhen. Die konträren Ergebnisse sind darauf zurückzuführen, dass für die Beurteilung der GCO ein kurzer Zeitraum von 12 Monaten berücksichtigt wird und Maßnahmen einen entscheidenden Einfluss in dieser Phase erbringen müssen. Darüber hinaus können die Studien für die Qualität der Finanzberichterstattung, gemessen anhand der Höhe der Accruals, verschiedene Befunde aufdecken. Während einerseits Belege für einen signifikanten und positiven Zusammenhang zwischen der Höhe der Accruals und der Wahrscheinlichkeit einer GCO vorhanden sind, sehen Butler et al. (2004) die Verwendung von Accruals zur Messung von Earnings Management kritisch, da die Beurteilung der Qualität der Finanzberichterstattung nicht zum Aufgabenbereich des Abschlussprüfers gehört.[227] Die Unterschiede resultieren aus den eingesetzten Modellen, die teilweise eine hohe Aussagekraft aufweisen, dann allerdings unter einer Fehlspezifikation leiden, weil eher Accruals mit Bezug zu einer finanziellen Notlage eines Unternehmens erfasst werden.

Das Audit Committee, welches zur Überwachung der Unternehmensführung eingesetzt wird, zeigt eine höhere Effektivität, wenn es mit unabhängigen Mitgliedern besetzt ist, die keine Beziehungen zum Unternehmen oder dem Management aufweisen. Ein unabhängigeres Audit Committee führt zu einer höheren Wahrscheinlichkeit, eine GCO ausgestellt zu bekommen, da abhängige Mitglieder eher im Sinne des Managements entscheiden und ihre Funktion als Überwachungsorgan vernachlässigen. Ist die Wirksamkeit der internen Kontrolle eingeschränkt, so kann das Management Going-Concern-Probleme eher verdecken und der Abschlussprüfer kann diese Themen nicht oder nur mit großem Aufwand identifizieren.

Werden die Faktoren beim Abschlussprüfer genauer beleuchtet, so lässt sich zunächst erkennen, dass Abschlussprüfer, die eine höhere Bildung, mehr berufliche Fortbildungskurse und eine längere Berufserfahrung vorweisen, die Going-Concern-Berichte übersichtlicher und präziser darstellen. Hinweise für einen negativen Zusammenhang zwischen der Erfahrung und der Wahrscheinlichkeit einer GCO erweisen sich allerdings als nicht signifikant. Eingeschränkt werden die Ergebnisse dadurch, dass sie zum Teil auf Typ II Fehlern basieren, die kein direktes Maß für die Qualität der Entscheidung des Abschlussprüfers über die Erteilung einer GCO sind. Die durch einen Typ II Fehler ex post festgestellte Fehlklassifikation ist nicht sicher mit einer fehlenden Normenkonformität der Prüfung gleichzusetzen. Vielmehr liegt bei der mit Unsicherheiten behafteten Beurteilung erst dann ein Prüferversagen vor, wenn der Prüfer sich gegen eine übliche Einschätzung entscheidet, obwohl sich an den

[227] Vgl. Butler et al. (2004, S. 140-142).

Rahmenbedingungen keine großen Veränderungen ergeben haben.[228] Die Ergebnisse für den Einfluss des Alters eines Abschlussprüfers sind, trotz des negativen Zusammenhangs zur Wahrscheinlichkeit der Ausstellung einer GCO, als nachrangig für die GCO-Entscheidung einzuschätzen. Hinsichtlich der Industriespezialisierung des Abschlussprüfers lassen sich keine eindeutigen Ergebnisse ermitteln. Einige Studien stellen einen positiven Zusammenhang zur Wahrscheinlichkeit einer GCO fest, was mit einem besseren Verständnis des Abschlussprüfers über die Sachverhalte und damit einer besseren Einschätzung der Going-Concern-Prämisse begründet werden kann. Gleichzeitig können andere Untersuchungen diesen Zusammenhang nicht bestätigen und finden nur insignifikante Resultate. Die unterschiedlichen Ergebnisse können dadurch bedingt sein, dass es verschiedene Definitionen der Industriespezialisierung gibt und die Spezialisierung nicht direkt beobachtbar bzw. messbar ist.

Ein wesentlicher Aspekt der diskutierten Faktoren ist die Unabhängigkeit des Abschlussprüfers in seiner Entscheidungsfindung. Bei der Betrachtung der finanziellen Abhängigkeit werden uneinheitliche Erkenntnisse gewonnen. Einerseits ist eine negative Verbindung zwischen den Nichtprüfungshonoraren und der Wahrscheinlichkeit einer GCO berechenbar, sodass von einer Einschränkung der Unabhängigkeit in der Urteilsbildung durch finanzielle Interessen ausgegangen werden kann. Andererseits präsentiert die Studie von Griffin und Lont (2010), dass regulatorische Maßnahmen, wie bspw. der SOX, den Zusammenhang abschwächen können.[229] Weitere Untersuchungen sehen zudem keinen signifikanten Zusammenhang und zeigen damit auf, dass die finanziellen Abhängigkeiten nicht vollständig im Vordergrund stehen, sondern auch andere Faktoren, wie z. B. die Reputation oder Risiken von Rechtsstreitigkeiten, Berücksichtigung finden. Dass die Erbringung von bestimmten zusätzlichen Dienstleistungen, in diesem Fall die Steuerberatung, sogar einen positiven Einfluss haben kann, zeigt Robinson (2008) in ihrer Studie.[230] Durch die weitere Tätigkeit bei einem Unternehmen können wichtige Informationen gewonnen und im Rahmen der Prüfung verwendet werden, sodass die Entscheidung über eine GCO auf Grundlage eines ausgedehnteren Kenntnisstandes erfolgen kann. Die Prüfungshonorare selbst sowie die relative Größe eines Mandanten für einen Abschlussprüfer liefern mehrheitlich keine signifikanten Zusammenhänge.

Ein differenziertes Bild ergibt sich bei der Größe des Abschlussprüfers. Während einige Untersuchungen zu dem Schluss kommen, dass Big-4-Prüfungsgesellschaften mit einer geringeren Wahrscheinlichkeit eine GCO erteilen, weil die geprüften Unternehmen größer sind und durch ihre bessere finanzielle Situation ein geringeres Risiko für eine Insolvenz aufweisen, gehen andere Studien von einem positiven Zusammenhang aus. Die größere Wahrscheinlichkeit einer GCO bei Big-4-Prüfungsgesellschaften wird mit einer höheren Prüfungsqualität im Vergleich zu Non-Big-4-Prüfungsgesellschaften sowie einer größeren Gefahr der Auswirkungen einer fehlerhaften Einschätzung der GCO auf die Reputation begründet.

Bei der Beziehung zwischen dem Abschlussprüfer und dem Mandanten sind vornehmlich die Beschäftigungsdauer des Abschlussprüfers und die Verzögerung der Berichterstattung von Bedeutung. Eine einheitliche Meinung zu der Wirkung der Beschäftigungsdauer kann nicht gefunden werden. Für die Beschäftigungsdauer des Prüfungspartners kann zwar eine negative Korrelation zur Wahrscheinlichkeit einer GCO ermittelt werden, was eine geringe Unabhängigkeit durch die lange Mandatsdauer begründen würde, dies ist jedoch nicht für die Prüfungsgesellschaft selbst gültig. Dem gegenüber zeigen weitere Untersuchungen eine größere Wahrscheinlichkeit eine GCO zu erhalten, wenn der Abschlussprüfer länger bei einem Unternehmen tätig ist. Ein denkbarer Grund hierfür wäre ein besseres Verständnis über das Unternehmen mit der Zeit, sodass der Abschlussprüfer seine GCO-Entscheidung zuverlässiger treffen kann. Andere Studien erkennen allerdings keine oder nur geringe Hinweise für eine Beziehung zwischen der Mandatsdauer und der Wahrscheinlichkeit einer GCO. Bei der Auswertung der Studien gilt es zu beachten, dass die Beendigung eines Mandats durch eine freiwillige oder verpflichtende Rotation induziert sein kann und die Untersuchungen aufgrund unterschiedlicher Rechtsvorgaben zu abweichenden Ergebnissen kommen. Ebenso ist die Identifizierung des Abschlussprüfers und damit die Ermittlung der Beschäftigungsdauer des Prüfungspartners u. U. nur eingeschränkt möglich, da die Offenlegung des Namens des Prüfungspartners nicht in jedem Land verpflichtend ist.[231] Des Weiteren muss die höhere Wahrscheinlichkeit eines Abschlussprüferwechsels bei finanziell gestressten Unternehmen berücksichtigt werden, da Unternehmen mit Problemen bei der Unternehmensfortführung eher zu einer kürzeren Mandatsdauer des Abschlussprüfers tendieren. Im Hinblick auf eine Verzögerung der Berichterstattung kann eine überwiegend positive Verbindung zur Wahrscheinlichkeit einer GCO konstatiert werden. Die längere Dauer für die Berichterstattung ist insbesondere durch einen höheren Zeitaufwand für weitergehende Prüfungshandlungen und durch die Verhandlungen über die identifizierten Probleme mit dem Management bedingt.

Wenn ein Unternehmen eine GCO erteilt bekommt, werden den Adressaten des Jahresabschlusses wichtige Informationen zur Unternehmensfortführung kommuniziert. Die Reaktion des Marktes hat dabei einen negativen Einfluss auf die Bewertung des Unternehmens, wie die Untersuchungen des Aktienkurses zeigen. Die bereits vor der eigentlichen Mitteilung der GCO erfolgte Informationsaufnahme und -verarbeitung und die damit verbundene geringere Reaktion um den Tag der Veröffentlichung aus früheren Studien, kann in späteren Studien nicht nachgewiesen werden.

[228] Vgl. Ruhnke und Frey (2015, S. 338-339).
[229] Vgl. Griffin und Lont (2010, S. 30).
[230] Vgl. Robinson (2008, S. 51-52).
[231] Vgl. ISA 700 (Revised) (2015, Tz. 46); Schlüter und Ratzinger-Sakel (2019, S. 8).

Die stärker negativen Auswirkungen einer GCO bei einem größeren Anteil an institutionellen Investoren legen Menon und Williams (2010) dar.[232] Dies resultiert aus dem besseren Verständnis sowie den verfügbaren Ressourcen institutioneller gegenüber nicht-professioneller Investoren. Zugleich kann eine Unterreaktion des Marktes auf die GCO für den Zeitraum von 12 Monaten nach Bekanntgabe ermittelt werden. Die Informationen einer GCO werden dementsprechend über einen längeren Zeitraum aufgenommen und verarbeitet und nicht direkt nach Bekanntwerden des Sachverhalts in den Aktienkurs eingepreist. Die Unterschiede zwischen den Studienergebnissen ergeben sich insbesondere durch Abweichungen bei dem betrachteten Zeitraum sowie dem Zeitpunkt, wann die Marktteilnehmer über eine GCO informiert werden.

Durch die vom Markt als negativ bewertete GCO kommt es beim Unternehmen selbst zu höheren Eigenkapitalkosten, da das Unternehmen ein höheres Risiko aufweist. Wie Chen et al. (2016) aufzeigen, wird der Kreditaufschlag nach einer GCO erhöht und neue Kredite werden mit einem geringeren Volumen sowie einer kürzeren Laufzeit vergeben.[233] Die GCO besitzt für die Stakeholder des Unternehmens somit aufgrund der signifikanten Marktreaktion einen wichtigen informativen Wert und kann z.B. Kreditverträge wesentlich beeinflussen. Allerdings können auch geringfügige Hinweise gefunden werden, dass die Kreditgeber ihre Entscheidung auf einer breiteren Informationsbasis treffen und dem modifizierten Prüfungsurteil eine weniger starke Bedeutung beimessen. Jedoch ist eine Generalisierbarkeit dieser Erkenntnis aufgrund der ausschließlichen Betrachtung von kleinen und mittleren Unternehmen in Finnland nicht gegeben.

Als weitere Informationsquelle können Ratingagenturen dienen. Eine Herabsetzung der Kreditwürdigkeit ist bei Unternehmen mit einer GCO wahrscheinlicher und führt zu einem Vertrauensverlust in das Unternehmen. Allerdings sind die Zusammenhänge noch nicht ausreichend erforscht und von der Struktur der Ratingagentur abhängig. Zudem sind die Informationen, die als Grundlage der Entscheidung der Ratingagenturen dienen, nicht ausreichend bekannt, weshalb die Ratingagenturen durch die fehlende Transparenz eine nachrangige Position einnehmen.

Die Theorie einer selbsterfüllenden Prophezeiung durch eine GCO wird in der Literatur überwiegend abgelehnt. Es können keine signifikanten Hinweise dafür gefunden werden, dass die Erteilung einer GCO bei Unternehmen zu einer höheren Wahrscheinlichkeit eines Zusammenbruchs führt. Weder in den USA noch in dem UK oder Australien führen die Studien zu verlässlichen signifikanten Ergebnissen. Dagegen sind die Ergebnisse, die in Belgien und Deutschland für eine selbsterfüllende Prophezeiung sprechen, wesentlich von der angewandten Forschungsmethode abhängig. Eine Veränderung der Stichprobe und des Matching-Verfahrens führt zu einer Widerlegung der vorherigen signifikanten Befunde. Des Weiteren ist zu berücksichtigen, dass eine GCO vorwiegend an Unternehmen mit finanziellen Schwierigkeiten ausgestellt wird. Dies muss in Untersuchungen berücksichtigt werden, da Unternehmen ohne finanzielle Schwierigkeiten wesentlich seltener eine GCO ausgestellt bekommen und bei diesen Unternehmen die Wahrscheinlichkeit für eine zukünftige Insolvenz naturgemäß geringer ist.

Die Konsequenz auf Seiten des Abschlussprüfers in Form eines Abschlussprüferwechsels bzw. bewussten Opinion Shoppings führt in der Gesamtbetrachtung der Untersuchungsergebnisse zu einer differenzierten Erkenntnis. Einerseits wird der Abschlussprüfer nach einer GCO mit einer größeren Wahrscheinlichkeit gewechselt, andererseits können die Studien keine Belege für ein erfolgreiches Opinion Shopping und einer damit verbundenen geringeren Wahrscheinlichkeit einer GCO nach dem Wechsel des Abschlussprüfers nachweisen. Werden in diesem Zusammenhang jedoch die Besetzung des Audit Committees und ein übermäßiges Selbstvertrauen des Managements integriert, so wird eine größere Wahrscheinlichkeit eines Abschlussprüferwechsels nach einer GCO deutlich, wenn das Audit Committee über einen größeren Anteil an Affiliated Direktoren verfügt bzw. das Management ein übermäßiges Selbstvertrauen aufweist. Hierzu kann von Carcello und Neal (2003) nachgewiesen werden, dass ein Opinion Shopping für ein uneingeschränktes Prüfungsurteil bei einem Audit Committee mit einem größeren Anteil an Affiliated Direktoren erfolgreich ist.[234]

Der Abschlussprüfer sieht sich durch den Verlust seines Mandanten mit finanziellen Einbußen konfrontiert, wobei die abgetretenen Prüfungshonorare bei einem Wechsel des Abschlussprüfers bei Unternehmen mit einer GCO wesentlich höher ausfallen als bei Unternehmen ohne einer GCO. Zudem kann sich eine GCO sowohl auf die Wahrnehmung des Abschlussprüfers als auch seine Reputation auswirken. Während Typ I Fehler eher positiv und als ein Signal für eine hohe Prüfungsqualität angesehen werden, stellen Typ II Fehler durch die fälschlicherweise Nicht-Ausstellung einer GCO einen negativen Anhaltspunkt dar und führen zu einer schlechteren Reputation beim Abschlussprüfer. Die Forschungsergebnisse zeigen außerdem Hinweise dafür, dass sich der Abschlussprüfer bei der Erteilung einer GCO mit einer geringeren Wahrscheinlichkeit mit Rechtsstreitigkeiten auseinanderzusetzen hat und finanzielle Zahlungen nach einem Urteil in der Regel geringer ausfallen.

6. Fazit

Die Analyse der Einflussfaktoren und Konsequenzen einer GCO zeigt auf, welche Determinanten aus wissenschaftlicher Sicht einen signifikanten Einfluss auf die Erteilung einer GCO haben und welche Folgen aus der Ausstellung einer GCO sowohl für die Unternehmen als auch die Abschlussprüfer resultieren. Bei Unternehmen in einer finanziell schwächeren Situation sowie einem unabhängigeren Audit Committee

[232] Vgl. Menon und Williams (2010, S. 2104).
[233] Vgl. Chen et al. (2016, S. 4-5).
[234] Vgl. Carcello und Neal (2003, S. 113).

und einer wirksamen internen Kontrolle kann ein positiver Zusammenhang zu der Wahrscheinlichkeit einer GCO nachgewiesen werden. Dagegen kann kein einheitliches Bild bei der Betrachtung der Industriespezialisierung, der finanziellen Abhängigkeit des Abschlussprüfers, der Größe des Abschlussprüfers und der Mandatsdauer gefunden werden. Während in einigen Studien ein signifikanter Zusammenhang bei diesen Faktoren ermittelt werden kann, ist bei anderen Untersuchungen eine insignifikante oder gegensätzliche Verbindung erkennbar.

Die Erteilung einer GCO führt im Allgemeinen zu einer negativen Marktreaktion, welche im Besonderen auch bei institutionellen Investoren festzustellen ist, woraus für das betroffene Unternehmen, wie dargestellt, höhere Kosten für die Kreditaufnahme entstehen und Kredite mit einem geringeren Volumen und einer kürzeren Laufzeit vergeben werden. Durch die Erteilung einer GCO werden außerdem den Marktteilnehmern zentrale Erkenntnisse über ein Unternehmen vermittelt. Eine selbsterfüllende Prophezeiung kann indes nicht bestätigt werden. Ebenso ist das Opinion Shopping durch Unternehmen als nicht erfolgreich anzusehen, obwohl die Wahrscheinlichkeit eines Abschlussprüferwechsels nach einer GCO größer ist. Für den Abschlussprüfer birgt vor allem die Nicht-Erteilung einer GCO bei einem Unternehmen, welches seine Tätigkeit zeitnah einstellen muss, Risiken in Form von einer Gefährdung der Reputation und durch mögliche Rechtsstreitigkeiten. Darüber hinaus muss der Abschlussprüfer bei einem Wechsel in Folge einer GCO Verluste von Prüfungshonoraren hinnehmen.

Anhand der Ergebnisse können wichtige Informationen für folgende Entscheidungen über die Regularien der Abschlussprüfung gewonnen werden. Zukünftige Studien können die einzelnen Faktoren und Konsequenzen einer GCO detaillierter und in Bezug auf die heute gültigen Rechtsvorschriften betrachten und alternative Forschungsansätze wählen.

Literatur

Abad, D., Sánchez-Ballesta, J. P. & Yagüe, J. (2017). Audit Opinions and Information Asymmetry in the Stock Market. *Accounting and Finance*, *57* (2), 565–595.

Abernathy, J. L., Guo, F., Kubick, T. R. & Masli, A. (2019). Financial Statement Footnote Readability and Corporate Audit Outcomes. *Auditing: A Journal of Practice & Theory*, *38* (2), 1–26.

Achyarsyah, P. (2016). The Analysis of The Influence of Financial Distress, Debt Default, Company Size, and Leverage on Going Concern Opinion. *Indian Journal of Applied Business and Economic Research*, *14* (10), 6767–6783.

Adam, S. (2007). *Das Going-Concern-Prinzip in der Jahresabschlussprüfung*. Wiesbaden, Springer.

Amin, K., Krishnan, J. & Yang, J. S. (2014). Going Concern Opinion and Cost of Equity. *Auditing: A Journal of Practice & Theory*, *33* (4), 1–39.

Averio, T. (2021). The Analysis of Influencing Factors on the Going Concern Audit Opinion – A Study in Manufacturing Firms in Indonesia. *Asian Journal of Accounting Research*, *6* (2), 152–164.

Banks, D. W. & Kinney, W. R. (1982). Loss Contingency Reports and Stock Prices: An Empirical Study. *Journal of Accounting Research*, *20*, 240–254.

Barth, M. E., Beaver, W. H. & Landsman, W. R. (1998). Relative Valuation Roles of Equity Book Value and Net Income as a Function of Financial Health. *Journal of Accounting and Economics*, *25* (1), 1–34.

Bartov, E., Gul, F. A. & Tsui, J. S. (2001). Discretionary-Accruals Models and Audit Qualifications. *Journal of Accounting and Economics*, *30* (3), 421–452.

Basioudis, I. G., Papakonstantinou, E. & Geiger, M. A. (2008). Audit Fees, Non-Audit Fees and Auditor Going-Concern Reporting Decisions in the United Kingdom. *ABACUS - A Journal of Accounting, Finance and Business Studies*, *44* (3), 284–309.

Behn, B. K., Kaplan, S. E. & Krumwiede, K. R. (2001). Further Evidence on the Auditor's Going-Concern Report: The Influence of Management Plans. *Auditing: A Journal of Practice & Theory*, *20* (1), 13–28.

Berglund, N. R. (2020). Do Client Bankruptcies Preceded by Clean Audit Opinions Damage Auditor Reputation? *Contemporary Accounting Research*, *37* (3), 1914–1951.

Berglund, N. R., Eshleman, J. D. & Guo, P. (2018). Auditor Size and Going Concern Reporting. *Auditing: A Journal of Practice & Theory*, *37* (2), 1–25.

Bills, K. L., Jeter, D. C. & Stein, S. E. (2015). Auditor Industry Specialization and Evidence of Cost Efficiencies in Homogenous Industries. *The Accounting Review*, *90* (5), 1721–1754.

Blay, A. D., Bryan, D. & Reynolds, J. K. (2016). The Going-Concern Market Anomaly Revisited. Zugriff auf https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2747774 (zuletzt abgerufen am: 23.03.2022)

Blay, A. D. & Geiger, M. A. (2001). Market Expectations for First-Time Going-Concern Recipients. *Journal of Accounting, Auditing & Finance*, *16* (3), 209–226.

Blay, A. D. & Geiger, M. A. (2013). Auditor Fees and Auditor Independence: Evidence from Going Concern Reporting Decisions. *Contemporary Accounting Research*, *30* (2), 579–606.

Blay, A. D., Geiger, M. A. & North, D. S. (2011). The Auditor's Going-Concern Opinion as a Communication of Risk. *Auditing: A Journal of Practice & Theory*, *30* (2), 77–102.

Brody, R. G. & Moscove, S. A. (1998). Mandatory Auditor Rotation. *National Public Accountant*, *43*, 32–35.

Brösel, G., Freichel, C., Toll, M. & Buchner, R. (2015). *Wirtschaftliches Prüfungswesen* (3. Aufl.). München.

Bruynseels, L. & Cardinaels, E. (2014). The Audit Committee: Management Watchdog or Personal Friend of the CEO? *The Accounting Review*, *89* (1), 113–145.

Bruynseels, L. & Willekens, M. (2012). The Effect of Strategic and Operating Turnaround Initiatives on Audit Reporting for Distressed Companies. *Accounting, Organizations and Society*, *37* (4), 223–241.

Butler, M., Leone, A. J. & Willenborg, M. (2004). An Empirical Analysis of Auditor Reporting and its Association with Abnormal Accruals. *Journal of Accounting and Economics*, *37* (2), 139–165.

Callaghan, J., Parkash, M. & Singhal, R. (2009). Going-Concern Audit Opinions and the Provision of Nonaudit Services: Implications for Auditor Independence of Bankrupt Firms. *Auditing: A Journal of Practice & Theory*, *28* (1), 153–169.

Carcello, J. V. & Neal, T. L. (2000). Audit Committee Composition and Auditor Reporting. *The Accounting Review*, *75* (4), 453–467.

Carcello, J. V. & Neal, T. L. (2003). Audit Committee Characteristics and Auditor Dismissals Following 'New' Going-Concern Reports. *The Accounting Review*, *78* (1), 95–117.

Carcello, J. V. & Palmrose, Z.-V. (1994). Auditor Litigation and Modified Reporting on Bankrupt Clients. *Journal of Accounting Research*, *32*, 1–30.

Carcello, J. V., Vanstraelen, A. & Willenborg, M. (2009). Rules Rather than Discretion in Audit Standards: Going-Concern Opinions in Belgium. *The Accounting Review*, *84* (5), 1395–1428.

Carey, P., Geiger, M. A. & O'connell, B. T. (2008). Costs Associated with Going-Concern-Modified Audit Opinions: An Analysis of the Australian Audit Market. *ABACUS - A Journal of Accounting, Finance and Business Studies*, *44* (1), 61–81.

Carey, P. & Simnett, R. (2006). Audit Partner Tenure and Audit Quality. *The Accounting Review*, *81* (3), 653–676.

Carson, E., Fargher, N. L., Geiger, M. A., Lennox, C. S., Raghunandan, K. & Willekens, M. (2013). Audit Reporting for Going-Concern Uncertainty: A Research Synthesis. *Auditing: A Journal of Practice & Theory*, *32* (Supplement 1), 353–384.

Carson, E., Fargher, N. L. & Zhang, Y. (2016). Trends in Auditor Reporting in Australia: A Synthesis and Opportunities for Research. *Australian Accounting Review*, *26* (3), 226–242.

Che, L., Langli, J. C. & Svanström, T. (2018). Education, Experience, and Audit Effort. *Auditing: A Journal of Practice & Theory*, *37* (3), 91–115.

Chen, P. F., He, S., Ma, Z. & Stice, D. (2016). The Information Role of Audit Opinions in Debt Contracting. Zugriff auf https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2599886 (zuletzt abgerufen am: 23.03.2022)

Chi, W., Myers, L. A., Omer, T. C. & Xie, H. (2017). The Effects of Audit Partner Pre-Client and Client-Specific Experience on Audit Quality and on Perceptions of Audit Quality. *Review of Accounting Studies*, *22* (1), 361–391.

Chow, C. W. & Rice, S. J. (1982). Qualified Audit Opinions and Auditor Switching. *The Accounting Review*, *57* (2), 326–335.

Christensen, B. E., Glover, S. M., Omer, T. C. & Shelley, M. K. (2016). Understanding Audit Quality: Insights from Audit Professionals and Investors. *Contemporary Accounting Research*, *33* (4), 1648–1684.

Citron, D. B. & Taffler, R. J. (2001). Ethical Behaviour in the U.K. Audit Profession: The Case of the Self-Fulfilling Prophecy Under Going-Concern Uncertainties. *Journal of Business Ethics*, *29* (4), 353–363.

Copley, P. A. & Doucet, M. S. (1993). Auditor Tenure, Fixed Fee Contracts, and the Supply of Substandard Single Audits. *Public Budgeting & Finance*, *13* (3), 23–35.

COVInsAG. (2021). *Gesetz zur vorübergehenden Aussetzung der Insolvenzantragspflicht und zur Begrenzung der Organhaftung bei einer durch die COVID-19-Pandemie bedingten Insolvenz (COVID-19-Insolvenzaussetzungsgesetz - COVInsAG) vom 27.03.2020, in der Fassung vom 15.02.2021*.

DeAngelo, L. E. (1981a). Auditor Independence, 'Low Balling', and Disclosure Regulation. *Journal of Accounting and Economics*, *3* (2), 113–127.

DeAngelo, L. E. (1981b). Auditor Size and Audit Quality. *Journal of Accounting and Economics*, *3* (3), 183–199.

DeFond, M. L., Lim, C. Y. & Zang, Y. (2016). Client Conservatism and Auditor-Client Contracting. *The Accounting Review*, *91* (1), 69–98.

DeFond, M. L., Raghunandan, K. & Subramanyam, K. R. (2002). Do Non-Audit Service Fees Impair Auditor Independence? Evidence From Going Concern Audit Opinions. *Journal of accounting research*, *40* (4), 1247–1274.

DeFond, M. L. & Zhang, J. (2014). A Review of Archival Auditing Research. *Journal of Accounting and Economics*, *58* (2-3), 275–326.

Dodd, P., Dopuch, N., Holthausen, R. W. & Leftwich, R. W. (1984). Qualified Audit Opinions and Stock Prices. *Journal of Accounting and Economics*, *6* (1), 3–38.

Dopuch, N., Holthausen, R. W. & Leftwich, R. W. (1986). Abnormal Stock Returns Associated with Media Disclosures of 'Subject to' Qualified

Audit Opinions. *Journal of Accounting and Economics*, *8* (2), 93–117.

Eickes, S. (2014). *Zum Grundsatz der Unternehmensfortführung in der Insolvenz*. Wiesbaden, Springer-Verlag. (zugl. Dissertation Universität des Saarlandes)

Elliott, J. A. (1982). 'Subject to' Audit Opinions and Abnormal Security Returns – Outcomes and Ambiguities. *Journal of Accounting Research*, *20* (2), 617–638.

Eutsler, J., Nickell, E. B. & Robb, S. W. G. (2016). Fraud Risk Awareness and the Likelihood of Audit Enforcement Action. *Accounting Horizons*, *30* (3), 379–392.

Fargher, N. L. & Wilkins, M. S. (1998). Evidence on Risk Changes Around Audit Qualification and Qualification Withdrawal Announcements. *Journal of Business Finance & Accounting*, *25* (7-8), 829–847.

Feldmann, D. & Read, W. J. (2013). Going-Concern Audit Opinions for Bankrupt Companies – Impact of Credit Rating. *Managerial Auditing Journal*, *28* (4), 345–363.

Firth, M. (2002). Auditor-Provided Consultancy Services and Their Associations with Audit Fees and Audit Opinions. *Journal of Business Finance & Accounting*, *29* (5-6), 661–693.

Francis, J., LaFond, R., Olsson, P. M. & Schipper, K. (2004). Costs of Equity and Earnings Attributes. *The Accounting Review*, *79* (4), 967–1010.

Francis, J. R. (2011). A Framework for Understanding and Researching Audit Quality. *Auditing: A Journal of Practice & Theory*, *30* (2), 125–152.

Francis, J. R. & Krishnan, J. (1999). Accounting Accruals and Auditor Reporting Conservatism. *Contemporary accounting research*, *16* (1), 135–165.

Francis, J. R. & Yu, M. D. (2009). Big 4 Office Size and Audit Quality. *The Accounting Review*, *84* (5), 1521–1552.

Frey, F. (2014). *Die self-fulfilling prophecy in der Abschlussprüfung – Eine empirische Analyse*. Freie Universität Berlin.

Gaeremynck, A. & Willekens, M. (2003). The Endogenous Relationship Between Audit-Report Type and Business Termination: Evidence on Private Firms in a Non-Litigious Environment. *Accounting and Business Research*, *33* (1), 65–79.

Gaver, J. J. & Utke, S. (2019). Audit Quality and Specialist Tenure. *The Accounting Review*, *94* (3), 113–147.

Geiger, M. A., Gold, A. & Wallage, P. (2021). *Auditor Going Concern Reporting: A Review of Global Research and Future Research Opportunities* (1. Aufl.). New York.

Geiger, M. A. & Raghunandan, K. (2001). Bankruptcies, Audit Reports, and the Reform Act. *Auditing: A Journal of Practice & Theory*, *20* (1), 187–195.

Geiger, M. A. & Raghunandan, K. (2002). Auditor Tenure and Auditor Reporting Failures. *Auditing: A Journal of Practice & Theory*, *21* (1), 67–78.

Geiger, M. A. & Rama, V. D. (2003). Audit Fees, Nonaudit Fees, and Auditor Reporting on Stressed Companies. *Auditing: A Journal of Practice & Theory*, *22* (2), 53–69.

Gerakos, J., Hahn, P. R., Kovrijnykh, A. & Zhou, F. (2016). Prediction Versus Inducement and the Informational Efficiency of Going Concern Opinions. Zugriff auf https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2802971 (zuletzt abgerufen am: 23.03.2022)

Goh, B. W., Krishnan, J. & Li, D. (2013). Auditor Reporting under Section 404: The Association between Internal Control and Going Concern Audit Opinions. *Contemporary Accounting Research*, *30* (3), 970–995.

Goodwin, J. & Wu, D. (2016). What is the Relationship Between Audit Partner Busyness and Audit Quality? *Contemporary Accounting Research*, *33* (1), 341–377.

Gramling, A. A., Krishnan, J. & Zhang, Y. (2011). Are PCAOB-Identified Audit Deficiencies Associated with a Change in Reporting Decisions of Triennially Inspected Audit Firms? *Auditing: A Journal of Practice & Theory*, *30* (3), 59–79.

Griffin, P. A. & Lont, D. H. (2010). Non-Audit Fees and Auditor Independence: New Evidence Based on Going Concern Opinions for U.S. Companies Under Stress. Zugriff auf https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1531523 (zuletzt abgerufen am: 23.03.2022)

Habib, A. (2013). A Meta-Analysis of the Determinants of Modified Audit Opinion Decisions. *Managerial Auditing Journal*, *28* (3), 184–216.

Hanlon, D., Khedmati, M. & Lim, E. K. (2019). Boardroom Backscratching and Audit Fees. *Auditing: A Journal of Practice & Theory*, *38* (2), 179–206.

Hardies, K., Breesch, D. & Branson, J. (2016). Do (Fe)Male Auditors Impair Audit Quality? Evidence from Going-Concern Opinions. *European Accounting Review*, *25* (1), 7–34.

Hardies, K., Vandenhaute, M.-L. & Breesch, D. (2018). An Analysis of Auditors' Going-Concern Reporting Accuracy in Private Firms. *Accounting Horizons*, *32* (4), 117–132.

Hassan, F. E. & Nasir, A. H. M. (2020). The Impacts of Investments in Internal Audit Functions on the Going Concern Opinion for Financially Distressed Firms. *Asian Journal of Accounting & Governance*, *14* (1), 25–36.

Heinrichs, M. (2018). *Going Concern Opinion, Managementkompetenz und Insolvenzwahrscheinlichkeit – Eine empirische Analyse*. Wiesbaden, Springer-Verlag.

HGB. (2020). *Handelsgesetzbuch vom 10.05.1897, in der Fassung vom 07.08.2021*.

Hopwood, W., McKeown, J. C. & Mutchler, J. F. (1994). A Reexamination of Auditor versus Model Accuracy within the Context of the Going-Concern Opinion Decision. *Contemporary Accounting Research*, *10* (2), 409–431.

IAS 1. (2020). *International Accounting Standard - Presentation of Financial Statements.*

IDW (Hrsg.). (2020). *Auswirkungen der Ausbreitung des Coronavirus auf die Rechnungslegung zum Stichtag 31.12.2019 und deren Prüfung, Fachlicher Hinweis des IDW.* Zugriff auf https://www.idw.de/blob/122498/31bce74e5b1413b91f74c9de1ea64383/down-corona-idw-fachlhinw-relepruefung-teil1-data.pdf (zuletzt abgerufen am: 23.03.2022)

IDW PS 270 n. F. (2021). *Die Beurteilung der Fortführung der Unternehmenstätigkeit im Rahmen der Abschlussprüfung.*

IDW PS 405. (2017). *Modifizierung des Prüfungsurteils im Bestätigungsvermerk.*

ISA 570 (Revised). (2015). *International Standard on Auditing 570 (Revised) - Going Concern.*

ISA 700 (Revised). (2015). *International Standard on Auditing 700 (Revised) - Forming an Opinion and Reporting on Financial Statements.*

Jiang, W., Rupley, K. H. & Wu, J. (2010). Internal Control Deficiencies and the Issuance of Going Concern Opinions. *Research in Accounting Regulation*, *22* (1), 40–46.

Jones, F. L. (1996). The Information Content of the Auditor's Going Concern Evaluation. *Journal of Accounting and Public Policy*, *15* (1), 1–27.

Kao, J. L., Li, Y. & Zhang, W. (2014). Did SOX Influence the Association Between Fee Dependence and Auditors' Propensity to Issue Going-Concern Opinions? *Auditing: A Journal of Practice & Theory*, *33* (2), 165–185.

Kaplan, S. E., Mowchan, M. J. & Weisbrod, E. H. (2014). Does Institutional Investor Behavior Influence the Market Reaction to Going Concern Audit Reports? Zugriff auf https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2433626 (zuletzt abgerufen am: 23.03.2022)

Kaplan, S. E. & Williams, D. D. (2013). Do Going Concern Audit Reports Protect Auditors from Litigation? A Simultaneous Equations Approach. *The Accounting Review*, *88* (1), 199–232.

Kausar, A. & Lennox, C. (2017). Balance Sheet Conservatism and Audit Reporting Conservatism. *Journal of Business Finance & Accounting*, *44* (7-8), 897–924.

Kausar, A., Taffler, R. J. & Tan, C. E. (2017). Legal Regimes and Investor Response to the Auditor's Going-Concern Opinion. *Journal of Accounting, Auditing & Finance*, *32* (1), 40–72.

Kida, T. (1980). An Investigation into Auditors' Continuity and Related Qualification Judgements. *Journal of Accounting Research*, *18* (2), 506–523.

Kim, M. H. J. (2021). Effects of Managerial Overconfidence and Ability on Going-Concern Decisions and Auditor Turnover. *Advances in Accounting*, *54* (1), 1–15.

Knechel, W. R. & Vanstraelen, A. (2007). The Relationship between Auditor Tenure and Audit Quality Implied by Going Concern Opinions. *Auditing: A Journal of Practice & Theory*, *26* (1), 113–131.

Koh, H. C. & Killough, L. N. (1990). The Use of Multiple Discriminant

Analysis in the Assessment of the Going-Concern Status of an Audit Client. *Journal of Business Finance & Accounting*, *17* (2), 179–192.

Krishnan, G. V. & Wang, C. (2015). The Relation Between Managerial Ability and Audit Fees and Going Concern Opinions. *Auditing: A Journal of Practice & Theory*, *34* (3), 139–160.

Krishnan, J. & Stephens, R. G. (1995). Evidence on Opinion Shopping from Audit Opinion Conservatism. *Journal of Accounting and Public Policy*, *14* (3), 179–201.

Leffson, U. (1988). *Wirtschaftsprüfung* (4. Aufl.). Wiesbaden, Springer-Verlag.

Lehmann, C. M. & Norman, C. S. (2006). The Effects of Experience on Complex Problem Representation and Judgment. *Behavioral Research in Accounting*, *18* (1), 65–83.

Lennox, C. S. (1999). Non-Audit Fees, Disclosure and Audit Quality. *European Accounting Review*, *8* (2), 239–252.

Li, C. (2009). Does Client Importance Affect Auditor Independence at the Office Level? Empirical Evidence from Going-Concern Opinions. *Contemporary Accounting Research*, *26* (1), 201–230.

Louwers, T. J., Messina, F. M. & Richard, M. D. (1999). The Auditor's Going-Concern Disclosure as a Self-Fulfilling Prophecy: A Discrete-Time Survival Analysis. *Decision sciences*, *30* (3), 805–824.

Marten, K.-U., Quick, R. & Ruhnke, K. (2020). *Wirtschaftsprüfung: Grundlagen des betriebswirtschaftlichen Prüfungswesens nach nationalen und internationalen Normen* (6. Aufl.). Stuttgart, Schäffer-Poeschel Verlag.

McKeown, J. C., Mutchler, J. F. & Hopwood, W. (1991). Toward an Explanation of Auditor Failure to Modify the Audit Opinions of Bankrupt Companies. *Auditing: A Journal of Practice & Theory*, *10*, 1–13.

Menon, K. & Williams, D. D. (2010). Investor Reaction to Going Concern Audit Reports. *The Accounting Review*, *85* (6), 2075–2105.

Meyer, Y. & Davids, K. (2018). Beurteilung der Fortführung der Unternehmenstätigkeit im Rahmen der Abschlussprüfung. *WPg – Die Wirtschaftsprüfung*, *71* (5), 263–271.

Mo, P. L. L., Rui, O. M. & Wu, X. (2015). Auditors' Going Concern Reporting in the Pre- and Post-Bankruptcy Law Eras: Chinese Affiliates of Big 4 Versus Local Auditors. *The International Journal of Accounting*, *50* (1), 1–30.

Mutchler, J. F., Hopwood, W. & McKeown, J. M. (1997). The Influence of Contrary Information and Mitigating Factors on Audit Opinion Decisions on Bankrupt Companies. *Journal of Accounting Research*, *35* (2), 295–310.

Niemi, L. & Sundgren, S. (2012). Are Modified Audit Opinions Related to the Availability of Credit? Evidence from Finnish SMEs. *European Accounting Review*, *21* (4), 767–796.

Nogler, G. E. (2004). Long-Term Effects of the Going Concern Opinion. *Managerial Auditing Journal*, *19* (5), 681–688.

Numan, W. & Willekens, M. (2012). Competitive Pressure, Audit Quality and Industry Specialization.. Zugriff auf https://www.ou.edu/dam/price/accounting/files/quality_competition_Oklahoma_May2012.pdf (zuletzt abgerufen am: 23.03.2022)

Parker, S., Peters, G. F. & Turetsky, H. F. (2005). Corporate Governance Factors and Auditor Going Concern Assessments. *Review of Accounting and Finance*, *4* (3), 5–29.

Pföhler, I., M./Knappe. (2021). Going Concern in Corona-Zeiten – Der Umgang mit den besonderen Unsicherheiten. *WPg – Die Wirtschaftsprüfung*, *74* (6), 354–360.

Raghunandan, K. & Rama, D. V. (1995). Audit Reports for Companies in Financial Distress: Before and After SAS No. 59. *Auditing: A Journal of Practice & Theory,*, *14* (1), 50–63.

Ratzinger-Sakel, N. V. (2013). Auditor Fees and Auditor Independence – Evidence from Going Concern Reporting Decisions in Germany. *Auditing: A Journal of Practice & Theory*, *32* (4), 129–168.

Read, W. J. & Yezegel, A. (2016). Auditor Tenure and Going Concern Opinions for Bank-rupt Clients: Additional Evidence. *Auditing: A Journal of Practice & Theory*, *35* (1), 163–179.

Reichelt, K. J. & Wang, D. (2010). National and Office-Specific Measures of Auditor Industry Expertise and Effects on Audit Quality. *Journal of Accounting Research*, *48* (3), 647–686.

Reynolds, J. K. & Francis, J. R. (2000). Does Size Matter? The Influence of Large Clients on Office-Level Auditor Reporting Decisions. *Journal of accounting and economics*, *30* (3), 375–400.

Robinson, D. (2008). Auditor Independence and Auditor-Provided Tax Service: Evidence from Going-Concern Audit Opinions Prior to Bankruptcy Filings. *Auditing: A Journal of Practice & Theory*, *27* (2), 31–54.

Ruhnke, K. (2003). Nutzen von Abschlussprüfungen: Bezugsrahmen und Einordnung empirischer Studien. *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung*, *55* (3), 250–280.

Ruhnke, K. & Frey, F. (2015). Einfluss der Finanzkrise auf das going concern-Berichterstattungsverhalten des Abschlussprüfers. Zugriff auf https://refubium.fu-berlin.de/bitstream/handle/fub188/21228/Ruhnke_Frey_BFuP-2015-3.pdf?sequence=1 (zuletzt abgerufen am: 23.03.2022)

Sarburg, M. & Mengwasser, M. (2017). Die Fortführung der Unternehmenstätigkeit im Lichte der Pflichten des Abschlusserstellers. *WPg – Die Wirtschaftsprüfung*, *70*, 20, 1179–1188.

Schlüter, J. K. & Ratzinger-Sakel, N. V. (2019). Audit Partners Signature Disclosure Effects on Non-Professional Investors' Credibility Assessments and Liability Perceptions. Zugriff auf https://ssrn.com/abstract=3442568 (zuletzt abgerufen am: 28.02.2022)

Simamora, R. A. & Hendarjatno, H. (2019). The Effects of Audit Client Tenure, Audit Lag, Opinion Shopping, Liquidity Ratio, and Leverage to the Going Concern Audit Opinion. *Asian Journal of Accounting Research*, *4* (1), 145–156.

Strickett, M., Hay, D. C. & Lau, D. (2021). The Going-Concern Opinion and the Adverse Credit Rating: An Analysis of Their Relationship. *Accounting Research Journal*. Zugriff auf https://www.emerald.com/insight/content/doi/10.1108/ARJ-04-2021-0135/full/html (zuletzt abgerufen am: 23.03.2022)

Subramanyam, K. & Wild, J. J. (1996). Going-Concern Status, Earnings Persistence, and Informativeness of Earning. *Contemporary Accounting Research*, *13* (1), 251–273.

Sundgren, S. & Svanström, T. (2014). Auditor-in-Charge Characteristics and Going-Concern-Reporting. *Contemporary Accounting Research*, *31* (2), 531–550.

Taffler, R. J., Lu, J. & Kausar, A. (2004). In Denial? Stock Market Underreaction To Going-Concern Audit Report Disclosures. *Journal of Accounting and Economics*, *38* (1), 263–296.

Vanstraelen, A. (2003). Going-Concern Opinions, Auditor Switching, and the Self-Fulfilling Prophecy Effect Examined in the Regulatory Context of Belgium. *Journal of Accounting, Auditing & Finance*, *18* (2), 231–254.

Wright, A. M. & Wright, S. (2014). Modification of the Audit Report: Mitigating Investor Attribution by Disclosing the Auditor's Judgement Process. *Behavioral Research in Accounting*, *26* (2), 35–50.

Wu, C. Y.-H., Hsu, H.-H. & Haslam, J. (2016). Audit Committees, Non-Audit Services, and Auditor Reporting Decisions Prior to Failure. *The British Accounting Review*, *48* (2), 240–256.

Ye, P., Carson, E. & Simnett, R. (2011). Threats to Auditor Independence: The Impact of Relationship and Economic Bonds. *Auditing: A Journal of Practice & Theory*, *30* (1), 121–148.

Junior Management Science 8(1) (2023) 123-147

Junior Management Science

journal homepage: www.jums.academy

Impact of Team Agility on Team Effectiveness: The Role of Shared Mental Models, Team Empowerment, and Team Reflexivity

Fabio Krüger

Technische Universität München

Abstract

While more and more organizations are adopting team agility as a new work approach to cope better with change, research still lacks a proper understanding of the human-side of team agility. To investigate the effectiveness and the human-side of team agility, this study builds on the IMO-framework. Team agility is investigated as the input factor, shared mental models (SMM) and team empowerment as mediators, team performance and team satisfaction as outcomes and team reflexivity as moderator between the relationships of mediators and outcomes. Data was collected from 23 agile working teams ($M_{team-size}$ = 3.48). Using linear regression both hypotheses, that team agility positively impacts SMM and team empowerment, were supported. This study found a significant total model effect for the relationship between team agility and team performance mediated by both, SMM and team empowerment. This study contributes to a better integration of the agile and teamwork literatures by identifying the roles of SMM and team empowerment on team effectiveness in an organizational context of team agility, as facilitating emergent team states.

Keywords: Agile work; team agility; team effectiveness; shared mental models; team empowerment.

1. Introduction

Increased development speed, shorter product lifecycles, demanding customers, the evolvement of online businesses, and many more characterizations of today's working environment lead to the need for organizational structures towards highly coordinated working teams (Wageman, Gardner, & Mortensen, 2012). To react appropriately to the rapidly changing environment and deal with increasing complexity, organizations tend to focus more on project teams instead of handling operations (Papadakis & Tsironis, 2018). To cope with these new arising challenges, organizations started adopting team agility (Conforto, Salum, Amaral, Da Silva, & De Almeida, 2014; Denning, 2018; Moe, Dingsøyr, & Dybå, 2010; Qumer & Henderson-Sellers, 2008; Serrador & Pinto, 2015; Tessem, 2014).

Team agility is a relatively new working approach to enable project teams a high degree of flexibility – the ability to change (Conforto, Amaral, da Silva, Di Felippo, & Kamikawachi, 2016). It emerged from the software development discipline to enhance the handling of requirements changes, realize productivity gains, and improve business alignments (Abrahamsson, Salo, Ronkainen, & Warsta, 2002; Beck et al., 2001; Campanelli & Parreiras, 2015). It is characterized by the ability to respond to change, extensive collaboration with customers, continuous improvement through team reflexivity of the project approach, and short work iterations (Beck et al., 2001). More and more executives from a wide range of business sectors are trying to implement team agility in their respective organizations (Denning, 2018; Tessem, 2014). But current research is mostly limited to either investigating a single or multiple specific agile methods, like Scrum or Extreme Programming (Campanelli & Parreiras, 2015; Schmidt, Kude, Heinzl, & Mithas, 2014; Serrador & Pinto, 2015), or focusing on the domain of software development teams (Dingsøyr, Fægri, Dybå, Haugset, & Lindsjørn, 2016; Williams & Cockburn, 2003). Agile methods are a part of team agility. They are designed to foster and enable the ability to change. Compared to team agility, they are limited to certain boundaries, like the context domain where they are applied. Thus, team agility is a more overarching approach to facilitate a higher flexibility of work teams in general. However, research currently lacks a proper theoretical grounding for explaining if and how team agility is related to team effectiveness. Especially how the human-side functions in such a work setting is still rather unclear. Therefore, understanding the impact

DOI: https://doi.org/10.5282/jums/v8i1pp123-147

of team agility on emergent states of teams can improve our knowledge of successful applications of team agility. In addition, it can extend the current state of team effectiveness research by providing team agility as a new input factor.

Intensive collaboration of all team members is standard in agile teams and emphasizes the importance of the humanside. One way to deal with the increased level of coordination and cooperation effort for teams is the development of shared knowledge structures regarding how to solve a given task and how to interact as a team (Mathieu, Heffner, Goodwin, Salas, & Cannon-Bowers, 2000). Similarity of knowledge team structures facilitates a shared understanding for upcoming tasks. This enables team members to better align and adapt their actions and behaviors to the task and the needs of the others (Cannon-Bowers, Salas, & Converse, 1993). In order to coordinate and adapt effectively, team members must anticipate the behavior of their teammates (Mathieu et al., 2000). This cognitive team mechanism refers to shared mental models (SMM). Team members use that to build an internal knowledge basis to manage their collaboration more effectively (Yu & Petter, 2014). The function of SMM is to enable team members to use their internal knowledge basis as a decision-making foundation for their future actions, thus, enabling better interaction (DeChurch & Mesmer-Magnus, 2010; Guchait & Hamilton, 2013; Mohammed, Ferzandi, & Hamilton, 2010).

Another way to cope with increased collaboration effort in teams is team empowerment. Team empowerment has been conceptualized along four dimensions by Seibert, Wang, and Courtright (2011), namely: competence, meaningfulness, self-determination, and impact. Research has shown that team empowerment is one lever to handle changing environmental and competitive demands. The ability to react fast and proactively (Campion, Medsker, & Higgs, 1993; Mathieu, Maynard, Rapp, & Gilson, 2008) through enhanced team empowerment can optimize both team performance and affective responses of team members (Kirkman & Rosen, 1999). Team agility is characterized by a self-organizing nature of the teams to enhance their flexibility (Hoda & Murugesan, 2016; Moe, Dingsøyr, & Dybå, 2008; Moe et al., 2010).

An essential component of team agility is the approach of continuous improvement through reflecting on the achieved outcomes after short performance periods (Beck et al., 2001; Moe et al., 2010). Taking a step back to reflect is also a well-known team process to gain, share, connect, and utilize knowledge (Argote, Gruenfeld, & Naquin, 1999; Schippers, Homan, & Van Knippenberg, 2013). The concept of team reflexivity refers to the extent to which teams discuss and modify the way they work to improve future collaboration (Otte, Knipfer, & Schippers, 2019; West, 1996). Thus, team reflexivity is two-fold: it is about assessing the past and using that assessment to plan the future (Konradt, Otte, Schippers, & Steenfatt, 2016). This team process affects the actions and behaviors of a team (Marks, Mathieu, & Zaccaro, 2001). Therefore, it is mostly applied by teams who face complex situations (e.g., changing environments) with difficult decisions (Schippers et al., 2013; West, 1996). Hence, team reflexivity can create conditions that impact the relationship of emergent states of a team, such as SMM or team empowerment, and its outcomes.

The goal of this study is to integrate the agile literature and teamwork research by investigating team agility as an essential driver for the emergence of team effectiveness. In order to theorize and empirically show the impact of team agility, I build on the Input-Mediator- Outcome (IMO) framework (Ilgen, Hollenbeck, Johnson, & Jundt, 2005) for team effectiveness (Mathieu et al., 2008). Team agility is examined as an input factor on team-level. The two team emergent states SMM (cognitive) and team empowerment (motivational) are explored as mediators. Team performance and team satisfaction are investigated as team outcomes. I use a quantitative empirical approach to compare cross-industry agile working teams without limiting them to a specific method.

Building on the theoretical foundation of teamwork research in combination with team agility as a potential new component of it, the present study provides three main contributions. First, this study examines the impact of team agility on team effectiveness independent of explicit agile methods or specific industries. In this way, I seek to generate a more holistic view of the value of team agility for teams across all industries. Second, my study contributes to the literature streams of agile and teamwork research by combining and integrating them. By doing so I aim to determine that team agility works as a driver for emergent states that facilitate team effectiveness. Further, I can respond to research calls for action from both areas. Stray, Moe, and Hoda (2018) called for more insights on how agile working teams can effectively coordinate their actions and behaviors as a response to internal and external influences. Also, Mathieu, Hollenbeck, van Knippenberg, and Ilgen (2017) asked for new understandings of teamwork in new work settings, such as the agile work context. By combining both theorized concepts, I seek to open a new perspective on team agility mechanisms in new organizational contexts. Third, I examine the effects of team reflexivity to create conditions in which the relationship between team emergent states and team outcomes can emerge. Team reflexivity is a specific team process that is an essential part of team agility (Beck et al., 2001). Team reflexivity is intended to create a high self-awareness in teams regarding their past actions and behaviors to enable a bright future. Thus, I aim to illuminate the potential moderating effect of team reflexivity to shed light on what we know about the impact of a fundamentally used work practice in agile working teams.

2. Theoretical Background and Hypotheses

The underlying framework for this research study is the IMO-framework established by Ilgen et al. (2005). It is a broadly established and used framework to explain team effectiveness components (e.g., performance) for various team settings (Mathieu et al., 2008). Thus, this study considers

the IMO-framework as a suitable approach to examine the impact of team agility on team effectiveness via the role of SMM and team empowerment. In the following, I summarize the core idea of the Input-Mediator-Outcome framework.

Inputs are defined as "antecedent factors that enable and constrain members' interactions" (Mathieu et al., 2008, p. 412). Input factors are categorized mainly into three different levels: varying from individual (e.g., characteristics) over team-level factors to organizational context factors (Mathieu et al., 2008). The level depends on their layer of deployment. Individual team member characteristics can be competencies like knowledge, skills, and abilities (KSAs). They can differ between individuals and teams, because teams face specific challenges (e.g., coordination) compared to individuals who work alone (Morgeson, Reider, & Campion, 2005). One team-level factor can be external leadership that assumes an impact on the whole team by the external leader, who is responsible for the performance of the team (Burke et al., 2006). Another team-level factor is the interdependence of a team that reflects the "extent to which team members cooperate and work interactively to complete tasks" (Stewart & Barrick, 2000, p. 137). An organizational and contextual input factor, for example, is openness climate (Mathieu et al., 2008). Openness climate refers to the ability of teams to communicate with each other in an open manner and, thus, exchange and share information with no boundaries (Mathieu, Gilson, & Ruddy, 2006). Team agility reflects one team-level input factor because it describes a working approach of a team. The application on team-level is characterized by a degree of use of specific ways of working. Thus, the team agility reflects the input factor in my study.

Mediators are variables that convey "important [...] influences with explanatory power for explaining variability in team performance and viability" (Ilgen et al., 2005, p. 520). They are mainly categorized into emergent states and team processes (Ilgen et al., 2005; Mathieu et al., 2008). Team processes are defined as the interdependent actions of team members that convert given inputs into outcomes (Marks et al., 2001). One example of a team process is coordination that encompasses the synchronization of various team members' interdependent actions (LePine, Piccolo, Jackson, Mathieu, & Saul, 2008). Emergent states are defined as "cognitive, motivational, and affective states of teams [that are] dynamic in nature and vary as function of team context, inputs, processes, and outcomes" (Marks et al., 2001, pp. 357–358). The main characteristic of emergent states is their development throughout the lifespan of a team. One affective emergent state, for example, is trust what is composed of the belief in a team's competencies and the safety that the individual or her/ his interests are protected (Ilgen et al., 2005; Mayer, Davis, & Schoorman, 1995). An example for a cognitive emergent state is strategic consensus, which represents the consensus on strategic priorities between team members (Kellermanns, Walter, Lechner, & Floyd, 2005). SMM and team empowerment reflect emergent states. SMM is a cognitive emergent state, thus, it refers to the creation of knowledge or understanding in teams based on their experiences or thoughts. Team empowerment is a motivational emergent state, thus, it refers to experiences of teams through feelings or emotions. SMM and team empowerment are the two mediators of my research study.

Outcomes are classified as the "results and by-products of team activity that are valued by one or more constituencies" (Mathieu et al., 2000, p. 273). They can either be measured based on quality or quantity, like team performance (Mathieu et al., 2006; Lester, Meglino, & Korsgaard, 2002), or members' affection and viability, like satisfaction or commitment (Balkundi & Harrison, 2006). Extant literature has examined team performance and satisfaction until today based on the IMO-framework for various team settings. I follow the literature by investigating team performance and team satisfaction as outputs.

The evidence has shown the reliability and validity of the underlying framework. It is used in various team settings to investigate overall team effectiveness by using different input factors, mediators, and outcomes (Ilgen et al., 2005; LePine et al., 2008; Mathieu et al., 2008; Maynard, Mathieu, Gilson, O'Boyle Jr, & Cigularov, 2013). Team agility as a relatively new phenomenon is particularly designed for employees who work in teams. Thus, team agility is recently gaining more attention due to the replacement or supplementation of traditional project management approaches (Laanti, Salo, & Abrahamsson, 2011). Even though the IMO-framework has been applied successfully in various team settings and with various types, empirical work using the IMO-framework within an agile setting is missing. Hence, the present study investigates team agility as an input factor, in line with the IMO-framework, and investigates its relationship with mediators and outcomes.

2.1. Team Agility

The concept of team agility has become widely adopted in team and management research. However, scholars remain unable to establish a common and unambiguous definition across all application domains (Conforto et al., 2016; Dybå & Dingsøyr, 2008; Mafakheri, Nasiri, & Mousavi, 2008; Sheffield & Lemétayer, 2013). This study follows the definition of Conforto et al. (2016, p. 667), denoting team agility as "the project team's ability to quickly change the project plan as a response to customer or stakeholders needs, market or technology demands to achieve better project and product performance in an innovative and dynamic project environment". I stick to the definition of Conforto et al. (2016) because it is primarily focused on the broader context of project management rather than limited to a focus of software development (Beck et al., 2001; Dybå & Dingsøyr, 2008; Mafakheri et al., 2008).

The "Agile Manifesto" is considered as emergence of team agility and states four main values[1]: (i) individuals and interaction over processes and tools; (ii) working software over comprehensive documentation; (iii) customer collaboration

[1] Originally developed for software development.

over contact negotiations; and (iv) responding to change over following a plan (Beck et al., 2001; Campanelli & Parreiras, 2015). It is supplemented by 12 underlying principles (Beck et al., 2001) that complete the definition. Team agility embodies four main components (Campanelli & Parreiras, 2015; Conboy & Fitzgerald, 2004) namely: "flexibility, velocity, learning and response to change, and leanness" (Ghezzi & Cavallo, 2020, p. 521).

Today, team agility is considered a new mainstay of project management to ensure long-term success (Denning, 2013). Hence, it has gained the interest of researchers in project team settings (Conforto et al., 2014; Conforto et al., 2016; Ghezzi & Cavallo, 2020; Misra, Kumar, & Kumar, 2009; Sheffield & Lemétayer, 2013). As traditional project management (TPM) approaches have been one of the main input factors of the IMO-Model to date, a clear distinction between TPM and team agility is helpful to explore team agility as a new input factor. This distinction supports a clear understanding of team agility management practices so that it can be investigated more generally instead of focusing on specific agile methods. This research paper follows Sheffield and Lemétayer (2013) suggestion to distinguish team agility from TPM according to three determinants: requirements uncertainty, customer collaboration, and team empowerment. Thus, team agility represents a new way of working together.

This study considers team agility as a team-level input factor of the IMO-framework. I examine team agility as a construct along six management practices. According to Conforto et al. (2014), those management practices are uniquely used in team agility and thus, differentiate it from TPM. This approach contrasts with traditional research about team agility, traditional research investigated team agility along specific agile methods (e.g., Scrum, pair programming, etc.) or agile practices (e.g., retrospective meetings, burndown charts). Those approaches neglect hybrid versions and hamper the comparability of teams across different industries (Abrahamsson et al., 2002; Dingsøyr et al., 2016; Dingsøyr, Falessi, & Power, 2019; Dybå & Dingsøyr, 2008; Guinan, Cooprider, & Faraj, 1998; Gustavsson, 2016; Hoda & Murugesan, 2016; Hummel, 2014; Moe et al., 2010; Serrador & Pinto, 2015). Building on the empirical support that team agility increases team effectiveness (Abrahamsson et al., 2002; Dingsøyr et al., 2016; Gustavsson, 2016; Schmidt et al., 2014; Serrador & Pinto, 2015), the present research focusses on team agility in terms of the management practices applied.

2.2. Shared Mental Models

SMM reflect a cognitive emergent state and thus, according to the IMO-framework, can represent a mediator (Ilgen et al., 2005; Mathieu et al., 2008) in the relationship between team agility and team performance and team satisfaction. From the perspective of multilevel theory, SMM is an emergent characteristic that stems from the knowledge and experience of individuals, but manifests as a collective characteristic (Klein & Kozlowski, 2000a). SMM are defined as "knowledge structures held by members of a team that enable them to form accurate explanations and expectations for the task, and in turn to coordinate their actions and adapt their behavior to demands of the task and other team members" (Cannon-Bowers et al., 1993, p. 228). SMM enable teams to adapt effectively through predicting what team members are planning next and which requirements need to be fulfilled to succeed (Cannon-Bowers et al., 1993). Some degree of similarity among the team members in terms of understanding the tasks and relationships within the team is necessary to accomplish team goals. Therefore, the concept of SMM explains the building of an internal knowledge basis of teams.

There are two different types of SMM that develop within a team. Researchers tend to categorize SMM in task-related SMM and team-related SMM (Cooke et al., 2003; Mathieu et al., 2000; Mohammed et al., 2010). First, task-related SMM include a shared understanding of work goals and performance requirements. Hence, it refers to the knowledge of how to execute tasks as a team. Typical tasks are knowing the goals, the interdependencies, and the process of how to approach a task. Second, team-related SMM include a shared understanding of interpersonal interactions, skills and responsibilities of other team members. Team-related SMM focus on shared understanding about how the team interacts, such as communication frequency, individual roles and responsibilities, and competencies (e.g., individual skills) of all team members (Cannon-Bowers et al., 1993).

Scholars have started to divide the evaluation of SMM development in teams into the dimensions of similarity and accuracy (Edwards, Day, Arthur Jr, & Bell, 2006; Mathieu, Heffner, Goodwin, Cannon-Bowers, & Salas, 2005; Mohammed et al., 2010). Similarity refers to the degree of internal consistency of team members' collective knowledge regarding teamwork and taskwork. Accuracy reflects the degree of SMM to "true state of the world" (Edwards et al., 2006, p. 728), which describes an objective view on teamwork and taskwork. This true score is often evaluated by experts. In the present study, I focus on the overall construct of SMM similarity. In what follows, I refer to a SMM similarity each time I mention SMM.

2.2.1. Team Agility on SMM

SMM are conceptualized to explain how teams are capable of dealing with complex and changing situations through different facets of knowledge, like declarative, procedural, and strategic knowledge (Marks et al., 2001; Mohammed et al., 2010). Developing SMM is similar to converging in the direction of a specific mindset. Aligned and adopted intra-team related guidelines can foster a team- and taskwork philosophy among team members. A philosophy is the core of the agile working's foundation. Abrahamsson et al. (2002) identified that working in an agile manner equals having a certain philosophy. That philosophy is captured by Beck et al. (2001) in the Agile Manifesto, which represents overarching guidelines for a successful application of team agility. It is further equipped with a set of methodologies and practices.

Together, agile working teams are provided with knowledge regarding their task- and teamwork.

For developing SMM various team interventions and team processes, like planning, reflexivity, leadership, and training, can act as an antecedent (Mohammed et al., 2010). Stout, Cannon-Bowers, Salas, and Milanovich (1999) have shown that effective planning results in higher SMM similarity. Effective planning allows team members to prioritize tasks, share information (or decide which kind of information should be shared), and discuss future events plus their implications for the team (e.g., how to back up each other). Orasanu (1994) identified that more planning contributes significantly to effective planning. An iterative planning approach is a fundamental agile project management practice, according to Conforto et al. (2014). Likewise, it is explicitly anchored in the twelve principles of "Agile Manifesto" (Beck et al., 2001). Moreover, the iterative planning is a major distinction between agile project management (Sheffield & Lemétayer, 2013) and a TPM that basically follows the initially set up plan (Boehm & Turner, 2005). Moreover, scholars (Klimoski & Mohammed, 1994) have identified various primary mechanisms that embrace a higher SMM similarity, such as communication or information sharing. According to the literature review of Gustavsson (2016), team agility enables teams to increase collaboration as well as improve knowledge sharing and better understanding of goals, tasks, and requirements. Practices of team agility, like iterative planning, frequent communication, and open information sharing, promote the development of SMM because they work like antecedents of SMM. Therefore, I assume a positive causal relationship between team agility and the SMM. Thus, I propose the first hypothesis:

H1a: Team agility positively impacts SMM.

2.3. Team Empowerment

Team empowerment reflects an motivational emergent state and thus, according to the IMO-framework, can represent a mediator (Ilgen et al., 2005; Mathieu et al., 2008) in the relationship between team agility and team performance and respectively team satisfaction. Team empowerment, according to Mathieu et al. (2006, p. 98), is defined as a team's "collective belief that they have the authority to control their proximal work environment and are responsible for their team's functioning". This definition encompasses four major domains of team empowerment: competence (to be able to perform tasks effectively), meaningfulness (in the sense of work), self-determination (to do task-related decisions), and impact (regarding outcomes (Kirkman & Rosen, 1999; Seibert et al., 2011). Team empowerment manifests among the team as shared perceptions of the empowerment level along the four dimensions (Chen, Kirkman, Kanfer, Allen, & Rosen, 2007).

2.3.1. Team Agility on Team Empowerment

Antecedents of team empowerment are encompassed in team agility, such as structural empowerment (Maynard et al., 2013) and high-performance management practices (Seibert et al., 2011). Antecedents of team empowerment enable its emergence. Structural empowerment reflects increased autonomy for employees through the transition of several activities to lower levels, such as scheduling or monitoring work. One characteristic of team agility is the high degree of self-organization and self-management of the teams (Moe et al., 2008). Thus, the underlying idea is that team agility gives teams increased autonomy in the structure of their work or task, which should improve their feeling of empowerment (Maynard et al., 2013). Chen, Sharma, Edinger, Shapiro, and Farh (2011) have shown that leadership distribution affects the embracement of team empowerment positively. Team agility similarly promotes shared leadership. It implements various roles (e.g., for Scrum: Product Owner, Scrum Master, Development Team) that share the decision-making power in order to empower the team (Cooper & Sommer, 2016). Thus, it should result in a greater feeling of project ownership and ultimately, enhance the feeling of autonomy.

High-performance management practices are reflected in "open information sharing, decentralization, participative decision making, extensive training, and contingent compensation" (Seibert et al., 2011, p. 983). According to Combs, Liu, Hall, and Ketchen (2006) a key component of empowering employees is information sharing. The kind of information differs between organizational roles and work tasks. The fundamental thought here is that team agility fulfills structural empowerment requirements and works with high-performance management practices. The nature of agile teams is described by a self-organizing and self-managing structure. Those features are anchored in the twelve principles of the Agile Manifesto (Beck et al., 2001) to ensure their autonomy. Thus, team agility advocates high degrees of autonomy and self-organized teams per se (Hoda, Noble, & Marshall, 2012; Moe et al., 2008). The team itself decides what tasks they approach, as well as when and how they approach them (Boehm & Turner, 2003; Dybå & Dingsøyr, 2008). One additional component of team agility is a high customer collaboration (Beck et al., 2001). The main goal of customer collaboration is to exchange and share information regarding expectations and updates. Team agility promotes those high- performance management practices that embrace team empowerment. Also, Gustavsson (2016) reported in his literature review that teams who worked agile increased their customer collaboration. Those findings are in line with Sharp and Robinson (2010), who go a little further and formulate the three C's of agile ways of working: collaboration, coordination, and communication. Combining the fundamentals of team agility, self-organized teams with high customer collaboration, I see reasonable ground for a positive causal relationship between team agility and team empowerment. Thus, I present the following hypothesis:

H1b: Team Agility positively impacts team empowerment.

2.3.2. Relationship of SMM and Team Empowerment

The two mediators SMM and team empowerment can also impact each other. SMM are intended to create a mutual understanding about taskwork and teamwork among team members (Klimoski & Mohammed, 1994). Additionally, they can improve collaboration, because it helps to share similar expectation concerning intra- and inter-team processes (Stout et al., 1999). Thus, it can foster the teams' feeling of impact because they create an internal knowledge basis that allows them to draw inferences about what they can achieve. Impact and competence are according to Seibert et al. (2011) two important dimensions of team empowerment. Further, SMM can enhance the feeling of competence within a team by promoting different roles to each team member such that the awareness about competencies can rise. A positive and vital role in the development of SMM is attributed to role differentiation (Cannon-Bowers, Salas, Blickensderfer, & Bowers, 1998) and performance monitoring (Rasker, Post, & Schraagen, 2000). Seibert et al. (2011) described role differentiation as well as performance monitoring in form high performance managerial practices as an antecedent of team empowerment. A high empowered team leads to a collectivistic belief of the team ability to successfully carry out certain tasks and foster a mutual understanding of the importance of that task (Seibert et al., 2011). Moreover, research has already stated that empowerment is related to a convergence of meanings and functional relations between individuals and on a team level (Chen et al., 2007). I see reasonable evidence for an interdependence between SMM and team empowerment. Thus, I formulate the following hypothesis:

> H2: SMM are positively related to team empowerment.

2.4. Team Performance

Team performance reflects one core team outcome (S. G. Cohen & Bailey, 1997). Thus, I investigate team performance as one outcome of the IMO-framework, which belongs to team- level outcomes (Ilgen et al., 2005; Mathieu et al., 2008) in this study. The main purpose of teams in organizations is to perform certain tasks (Ilgen, 1999). Thus, the ultimate goal of teams is considered to be a great team performance (Bommer, Johnson, Rich, Podsakoff, & MacKenzie, 1995). There are multiple ways to capture the performance of teams. The measurement of team performance mostly depends on the context. According to Mathieu et al. (2008) these context domains can vary from an organizational performance over team performances (e.g., team performance behaviors) to team member performance (e.g., affect and viability). The present study focuses on team performance outcomes as it represents the quality resulting from team- and taskwork processes (Klein & Kozlowski, 2000a).

2.4.1. SMM on Team Performance

The impact of SMM on team performance has been investigated in several manners regarding the work dimensions (team-related and task-related work) and the evaluation dimensions (similarity and accuracy). DeChurch and Mesmer-Magnus (2010) have shown in their meta-analysis that SMM is a significant predictor of team performance. Those findings are in line with earlier studies (Lim & Klein, 2006; Mathieu et al., 2005), which have shown that task-related as well as team-related SMM similarity were a good predictor of team performance. A more recent study (Guchait & Hamilton, 2013) showed also that team learning behaviors were positively related to the development of SMM which, in turn, had a significant effect on team performance. SMM enable a team to understand and process (new) information more consistently, explain situations equally, and predict future events (Mohammed et al., 2010). Thus, SMM allow a better task coordination and behavioral adjustments. These improve the decision making and the performance of a team (Cannon-Bowers et al., 1993).

Mohammed et al. (2010) concluded in their meta-analysis that a positive relationship between SMM similarity and team performance was found. Various scholars have presented a critical mass of studies under various conditions with different SMM measurements supporting the conclusion of Mohammed et al. (Cannon-Bowers et al., 1993; Cooke et al., 2003; Ensley & Pearce, 2001; Lim & Klein, 2006; Mathieu et al., 2000; Mathieu et al., 2005; Maynard & Gilson, 2014; Stout et al., 1999). SMM help teams to better understand the events in their environment in a threefold manner. First, they manifest a similar understanding of information among team members. Second, teams generate a shared explanation of events in their environment. Third, teams can more precisely predict future events based on this shared understanding (Klimoski & Mohammed, 1994; Mohammed et al., 2010). Thus, I assume that this results in increased quality of team performance. Based on this body of existing research, the following hypothesis is proposed:

> H3a: SMM positively impact team performance.

2.4.2. Team Empowerment on Team Performance

Team empowerment has been shown to be a good predictor of team performance (Chen et al., 2007; Kirkman, Rosen, Tesluk, & Gibson, 2004; Seibert et al., 2011). Seibert, Silver, and Randolph (2004) have shown a significant impact of team empowerment on individual performance and job satisfaction, Chen et al. (2007) supported those findings by showing that team empowerment is positively related to leadership climate and team performance. Jiang, Flores, Leelawong, and Manz (2016) found that team empowerment predicts knowledge sharing and thus, fosters team performance.

Seibert et al. (2011) have presented evidence that team performance is a behavioral consequence of team empowerment (Chen et al., 2007; Kirkman et al., 2004; Seibert et al., 2004; Srivastava, Bartol, & Locke, 2006). An empowered team develops a higher sense of ownership based on improved initiative opportunities of team members (Spreitzer, Noble, Mishra, & Cooke, 1999). Empowered teams can respond quickly to varying task demands by better leveraging

their capabilities, interests, and availability (Maynard et al., 2013). By optimizing the overall capabilities within the team and better distributing tasks and responsibilities, team performance should increase. Therefore, the following hypothesis is proposed:

> H3b: Team empowerment positively impacts team performance.

2.5. Team Satisfaction

Team satisfaction reflects one core team outcome (S. G. Cohen & Bailey, 1997). Thus, I investigate team satisfaction as another team-level outcome of the IMO-framework (Ilgen et al., 2005; Mathieu et al., 2008) in this study. Team satisfaction is the aggregated value of individual job satisfaction among team members. Job satisfaction is a multifaced concept that can be broken down to "how people feel about their jobs and different aspects of their jobs [...] that is, whether or not the job met the employee's physical and psychological needs for the things provided by work" (Spector, 1997, p. 2).

2.5.1. SMM on Team Satisfaction

The impact of SMM on team satisfaction has remained relatively unexplored by literature (Chou, Wang, Wang, Huang, & Cheng, 2008; Mohammed et al., 2010). SMM facilitate high-quality interpersonal relationships that contribute positively to the job satisfaction of each individual team member. SMM enable team members to perceive their surrounding environment in a more similar manner (Mathieu et al., 2000). This similarity fosters a team attitude of cooperation and assistance (Chou et al., 2008). This attitude is reflected in the anticipation of each other's needs, which, in turn, establishes high-quality interpersonal relationships in a team. Those relationships lead to a more positive feeling of each team member at work due to a better interaction in the team. Empirical evidence has shown that a positive relationship between a positive feeling at work and job satisfaction exists (Bretz & Judge, 1994). Hence, I suggest the following hypothesis:

> H4a: SMM positively impact team satisfaction.

2.5.2. Team Empowerment on Team Satisfaction

The impact of team empowerment on team satisfaction has rarely been investigated. The main focus of research has been on job satisfaction (Seibert et al., 2011). As team satisfaction is the aggregated value of job satisfaction, I argue that team empowerment positively impacts job satisfaction and hence, team satisfaction. Team empowerment is greater when individuals experience meaningfulness in their work, which contributes to increased job satisfaction (Seibert et al., 2004). Meaningfulness is a core component of team empowerment (Spreitzer, 1995). Also, evidence showed that the stimulation of meaningfulness is the best predictor of job satisfaction compared to the other three components of empowerment (Spreitzer, Kizilos, & Nason, 1997). Those findings are in line with earlier work regarding job satisfaction. Locke (1976) showed that individual fulfilment of values because of one's own work leads to job satisfaction. Some scholars (Kirkman & Rosen, 1999; Seibert et al., 2004) validated this and presented similar findings in additional studies. Thus, I argue that an empowered team experiences a great feeling of meaningfulness and that feeling leads to greater team satisfaction. Team empowerment can also lead to a feeling of responsibility and ownership of work (Maynard et al., 2013). Team Empowerment is likely to facilitate the pursuit of continuous improvement in work process and innovative solutions. Therefore, an empowered team strives for better outcomes that, in turn, result in increased efficiency and productivity (Seibert et al., 2011). Some scholars have shown that team empowerment supports the fulfillment of individuals' needs at work through increasing meaningfulness or self-determination (Deci & Ryan, 1985; Hackman & Oldham, 1980). The fulfillment of individual needs at work are related to satisfaction. The more the needs are fulfilled, the more satisfied individuals are (Locke, 1976). Thus, I argue that team empowerment enables a feeling of autonomy which satisfies an individual need at work. In doing so, team empowerment becomes an antecedent of team satisfaction. (Seibert et al., 2011). Hence, I propose the following hypothesis:

> H4b: Team empowerment positively impacts team satisfaction.

2.6. The Mediational Role of SMM and Team Empowerment

As explained previously, I hypothesize that team agility positively impacts SMM and team empowerment. Based on an extant body of literature (Gustavsson, 2016; Mathieu et al., 2000; Maynard et al., 2013; Mohammed et al., 2010; Seibert et al., 2011), this study also postulates a positive relationship between each mediator (SMM and team empowerment) and each outcome (team performance and satisfaction). Hence, this study argues that team agility positively impacts SMM and team empowerment, which, in turn, positively impact team performance and team satisfaction. Thus, I propose the following:

> H5a: SMM mediate the positive relationship between team agility and team performance.
>
> H5b: SMM mediate the positive relationship between team agility and team satisfaction.
>
> H6a: Team empowerment mediates the positive relationship between agility and team performance.
>
> H6b: Team empowerment mediates the positive relationship between agility and team satisfaction.

2.7. Team Reflexivity

Team reflexivity refers to, based on the proposed taxonomy by Marks et al. (2001), a transition processes and is defined as "the extent to which team members collectively reflect upon the team's objectives, strategies and processes and

adapt them to current and anticipated endogenous or environmental circumstances" (West, 1996, 559). Hence, team reflexivity encompasses a dual focus: first, to reflect on and interpret the progress up to now and second, to align on future actions to ensure future accomplishments (LePine et al., 2008). It can be subsumed to inspect the current work context (e.g., goals or strategy) and adapt the working methods to ensure an effective response. Team reflexivity allows team members to communicate openly and frequently over the past working period. Thus, there is the opportunity to talk about what went well and where is still room for improvement. Teams can either conduct a self-regulated or a guided reflexivity session. The two options mainly differ in their initiation. Self-regulated refers to a more autonomous and spontaneous approach whereas guided refers to a more organized and formalized session (Otte et al., 2019). In what follows, I refer to a guided team reflexivity session each time I refer to team reflexivity.

2.7.1. The moderating role of team reflexivity on the relationship between SMM and team performance

Team reflexivity is a transition process (Marks et al., 2001) that aims to collectively discuss how past team activities have or have not contributed to the status quo of the team (Otte et al., 2019). Some specific characteristics of team reflexivity include understanding, conversation, preparing, as well as transitioning to action, of group task facets (Schippers et al., 2013). It occurs mainly in transition phases when teams assess recent activities and plan next steps. Therefore, team reflexivity is typically conducted at the end of one performance period (Marks et al., 2001). Hence, team reflexivity promotes an atmosphere to share and integrate the experiences of all team members. This atmosphere enables open communication and information sharing that contribute to high self-awareness regarding the tasks and interaction in a team (Schippers et al., 2013). In contrast, non-reflective teams lack in their degree of self- awareness of their tasks and interactions. A highly reflective team tends to constantly observe its work environment. Such an assessment supports the adaptability to changing conditions of a team (West, 2012). This work climate can facilitate shared cognitive knowledge among team members and also work as motivation to improve the team performance. Thus, team reflexivity creates conditions in which a shared understanding of team tasks and team interactions can rise and have a stronger positive impact on team performance. Accordingly, I assume that team reflexivity can moderate the relationship between SMM (a cognitive emergent state) and team performance such that different levels of team reflexivity can lead to an increased/decreased relationship between SMM and team performance. I, therefore, propose:

> H7a: Team reflexivity moderates the positive relationship between SMM and team performance in such a way that the relationship is stronger when team reflexivity is high.

2.7.2. The moderating role of team reflexivity on the relationship between SMM and team satisfaction

Team reflexivity incorporates discussions about recent strategies and team processes, like communication (Schippers et al., 2013). Team reflexivity has a dual focus: interpreting the past and planning for the future. Thereby, it can generate working conditions in which team members share good and bad experiences from the last performance period. Further, it is not limited to the performance dimensions: teams also reflect about their behavior (Konradt et al., 2016). Team reflexivity encourages teams to agree on certain guidelines under which they will operate in the future. The opportunity to adjust those guidelines supports conditions in which shared perspectives can rise and team members can feel more understood. Employees pursue fulfillment at work that impacts their satisfaction (Al Jenaibi, 2010). The ability to engage in a pleasant work behavior supports that satisfaction (Mullins, 2016). Such an environment encourages the sharing of experiences and agreement on a satisfactory common way of working. Thus, team reflexivity creates conditions in which a shared understanding of team tasks and team interactions can rise and have a stronger positive impact on team satisfaction. Accordingly, I assume that team reflexivity can moderate the relationship between SMM and team satisfaction such that different levels of team reflexivity can lead to an increased/decreased relationship between SMM and team satisfaction. I, therefore, propose:

> H7b: Team Reflexivity moderates the positive relationship between SMM and team satisfaction in such a way that the relationship is stronger when team reflexivity is high.

2.7.3. The moderating role of team reflexivity on the relationship between team empowerment and team performance

Team reflexivity enables a team to cope with internal and external circumstances through reviewing the status quo and provide ideas for the future (Hoegl & Parboteeah, 2006). Team reflexivity refers to a mutual session of the team and is characterized by frequent communication about what went well and where there is still room for improvement (Tjosvold, Tang, & West, 2004). By doing so, the process of team reflexivity enables a team to exchange knowledge and to foster a feeling of competence, meaningfulness, responsibility, and control. By clearing up misunderstandings and addressing explicit wishes for future cooperation, the interaction with each other can be further improved (Schippers et al., 2013). A team with a higher degree of reflexivity is aware of its strengths and weaknesses and should emphasize their expertise to generate an atmosphere with a feeling of competence. Reviewing the progress and adjusting methods to ensure future success are drivers for a greater feeling of competence, meaningfulness, responsibility, and control. Team reflexivity enables a team to better understand how they collaborate (Tjosvold et al., 2004). Thereby, team reflexivity allows teams who feel responsible for their functioning

to reach greater outcomes and feel more satisfied. Accordingly, I assume that team reflexivity can moderate the relationship between team empowerment and team performance such that different levels of team reflexivity can lead to an increased/decreased relationship between team empowerment and team performance. I, therefore, propose:

> H7c: Team reflexivity moderates the positive relationship between team empowerment and team performance in such a way that the relationship is stronger when team reflexivity is high.

2.7.4. The moderating role of team reflexivity on the relationship between team empowerment and team satisfaction

The inspect and adapt approach of team reflexivity allows a team to assess their current situation (Otte et al., 2019). To take a step back and critically reflect on the achieved results allows a team to share and evaluate information (Otte, Konradt, Garbers, & Schippers, 2017). Moreover, they can consider their work from a new perspective with all its facets, like the impact for their organization or the meaningfulness to themselves. This provides a working environment in which team members become much more aware of what they have accomplished (Konradt, Schippers, Garbers, & Steenfatt, 2015). Further, they can adjust their work towards a more fulfilling direction. Team reflexivity is key component of employees' well-being (West, 2012). The satisfaction of an employee is related to its perception of the work environment. A good-natured wok environment can contribute to its satisfaction (Mullins, 2016). Reviewing the progress and adjusting methods to ensure future success are the drivers for a greater feeling of competence, meaningfulness, responsibility, and control. Team reflexivity enables a team to better understand how they collaborate (Tjosvold et al., 2004). Thereby, team reflexivity allows teams who feel responsible for their functioning to reach greater outcomes and feel more satisfied. Accordingly, I assume that team reflexivity can moderate the relationship between team empowerment and team satisfaction such that different levels of team reflexivity can lead to an increased/decreased relationship between team empowerment and team satisfaction. I, therefore, propose:

> H7d: Team reflexivity moderates the positive relationship between team empowerment and team satisfaction in such a way that the relationship is stronger when team reflexivity is high.

3. Method

To explore the impact of team agility on the emergence of SMM and team empowerment and, as a result, on team performance and team satisfaction, I conducted an online survey.

3.1. Sample

Twenty-three teams ($M_{team-size} = 3.48$ individuals per team, $MIN = 3$, $MAX = 6$) participated in my online survey. Every participant answered the same questionnaire regarding their project team (see below for questionnaire details). The final sample after applying predefined exclusion criteria consisted of 80 individuals of twenty-three teams. They were rewarded by receiving the results once the research study was finished. 44 participants indicated that they wanted to be informed about the results (either by receiving a management summary or by getting invited to a webinar). In total, the participants were mostly male (58.8%) and a bit less female (40.0%) and a few diverse (1.2%) with an average age of $M_{age} = 30.85$ years, ($SD_{age} = 7.28$). The sample consisted of 13 different nationalities who had worked in 14 different industries. The sample showed diverse functional roles. The participants showed a wide range of company sizes (number of employees and revenue) in which they were employed (see Appendix A). The sample consisted only of participants with experience of agile working practices from different industries, different functional departments, and different organizational roles. Those differences helped to investigate the phenomenon of team agility on a broader scale.

Before using the sample, I applied data cleaning and data preparation practices for analysis. I screened the data and excluded unusable data following predefined criteria. First, I only considered completed surveys. 99 teams ($n_{individuals} = 424$) started the survey. 43 teams ($n_{individuals} = 116$) completed the survey with a response rate of 43.43% (individuals = 27.36%). Followed by the exclusion of teams with less than three participants. Three participants per team ensure a representative team perspective. This led to 26 teams in total ($n_{individuals} = 94$). Further, two individuals were screened out automatically because they said that they did not work with agile methods or practices. This did not lead to an exclusion of an additional team due to sufficient remaining members per team of those 26 teams ($n_{individuals} = 92$). In addition, three teams were excluded due to invariant responses ($n_{individuals} = 80$). Invariant responses are defined as repeatedly using the same answer option (DeSimone, Harms, & DeSimone, 2015). As a cutoff-value, I followed the conventional approach by Huang, Curran, Keeney, Poposki, and DeShon (2012) of 14.

Furthermore, I also checked for exclusion cases of very short completion time (below 5 minutes) and unreliable responses (Goldammer, Annen, Stöckli, & Jonas, 2020) due to their psychometric synonyms with a recommend cutoff-value of 0.03 (DeSimone et al., 2015; Huang et al., 2012; Johnson, 2005). These cases did not lead to an exclusion of a participant or team. Thus, it is not shown in the Figure 2. The respective flowchart illustrates all previously explained exclusion cases.

3.2. Design and Procedures

3.2.1. Research Design

The purpose of this research study was to explore the impact of agile ways of working on team effectiveness and to

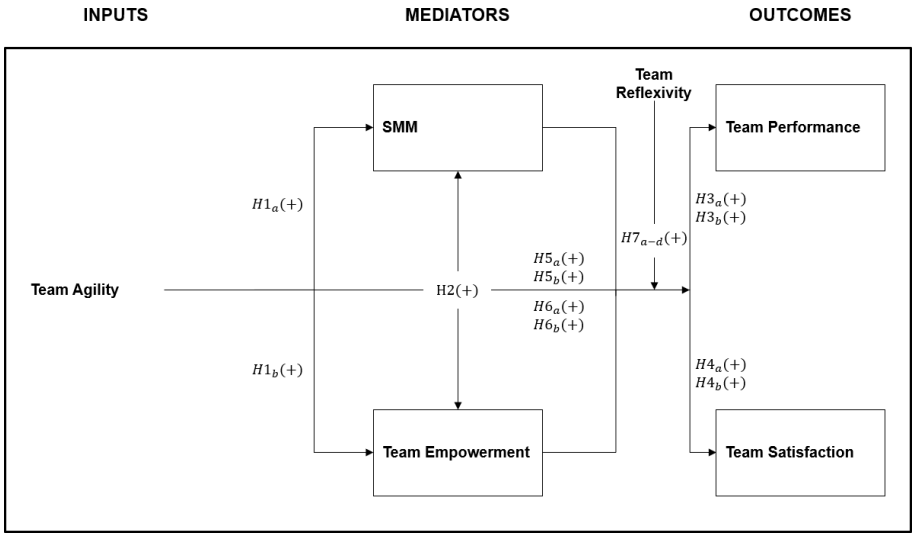

Figure 1: Research model.

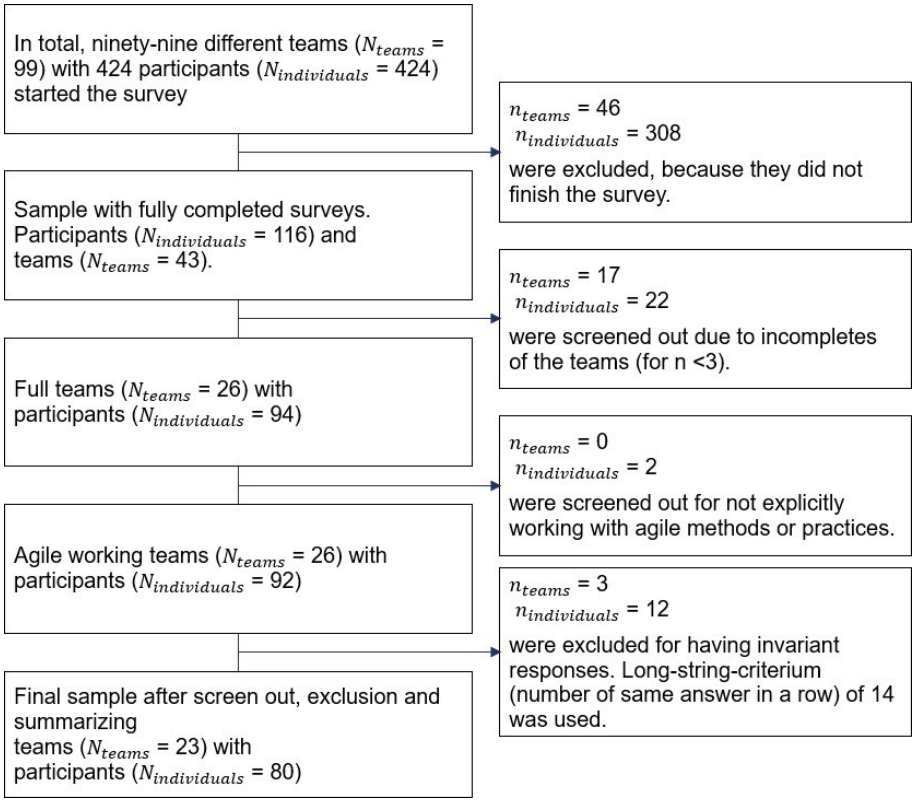

Figure 2: Flow Chart of Case Exclusions.

understand the role of team processes (SMM and team empowerment) in this relationship. Thus, I developed 16 hypotheses to test my assumptions. To validate or falsify these assumptions, I chose a cross-sectional, quantitative study design in form of an online survey. An online survey as a research design represents an appropriate instrument to collect data from a large number of individuals and to draw inferences regarding a larger group (Marshall & Rossmann, 2006). One strong advantage of online surveys is the capability to collect data fast and easily. Further, the distribution channels of an online survey allowed me to gather information from a more comprehensive participant sample (Döring & Bortz, 2016). Another advantage is the opportunity to apply appropriate analysis methods for online data (Laanti et al., 2011).

3.2.2. Procedure

To prepare and distribute the online questionnaire, Unipark[2], a web-based survey software program, was used. Prior to distributing the survey, I pretested it. The pilot survey was sent to some colleagues of mine who had experience in working with agile methods. They were asked to provide feedback regarding the understanding of the self-developed team agility construct, the approximated length of the survey, and general errors (e.g., spelling mistakes or incorrect formatting). The goal was to optimize the survey to avoid potential dropouts based on their feedback. Hence, the item "My team uses the 'product vision' concept." was changed to "My team uses the 'product vision' concept (e.g., north star)." in order to be clearer. The pilot survey also provided an approximation for the average time to complete the survey (Creswell, 2008), which was relatively long. Therefore, the number of open-questions regarding the use of agile ways of working on an organizational-level was reduced from six to two. Moreover, the pilot survey highlighted spelling mistakes and items that were difficult to understand that were then changed accordingly.

I created 120 specified team links. Thus, the links helped to first, categorize single participants to one team and second, keep track of the progress of the teams. Only people that used agile practices were contacted. I followed three different approaches to contact them. First, I contacted personal acquaintances directly via LinkedIn[3]. Second, I contacted people that I did not know personally via LinkedIn by searching for keywords in profiles like "AGILE", "AGILE TRANSFORMATION", "SCRUM", "SCRUM MASTER", and "PRODUCT OWNER". I tried to contact those people by sending a request to connect. In these cases, I sent a personal message to each one. I explained that I was conducting a survey as part of a research study. Further, I introduced the purpose of the study and asked whether they have worked with agile methods (or parts of it). I offered to share an excerpt of the results[4] with them once the study was finished. That offering had two intentions: First, to increase the willingness to participate (Döring & Bortz, 2016) and second, to increase the validity of self-reporting measures through honest responses (Mertens, 2005). The message ended with the question of if they would like to participate or if there are still open questions (see Appendix B). After receiving a positive response, I sent a follow-up message that included a specified team link to the survey. Further, I added information regarding the approximate length of the study (10 minutes) and that two-language options (English and German) available (see Appendix C). At the end, I reminded the participants to distribute the link to the survey by themselves in their team.

The team-specific link enabled me to track how many participants per team completed the online survey at any point of time. If after two weeks, less than three members of a team had participated, I sent another reminder message via LinkedIn (see Appendix D). If one week later, the minimum of three participants had still not been reached, I sent another reminder message. After that, no further reminders followed. The follow-up reminder emails were sent in order to encourage participation (Dillman, Smyth, & Christian, 2014).

The third approach to collect data differs to the previous ones. I joined eight groups on LinkedIn and Xing[5] to extend my reach. I searched for groups with one of the following keywords "AGILE", "AGILE TRANSFORMATION", "SCRUM". In each group, I created one identical forum post (see Appendix E) explaining the research study according to the personal messages (approach one and two). Moreover, the post contained a universal link to the study. The universal link differs in such a way that I was not able to send any reminders, because I received no information about who participated and when.

3.2.3. Study Design

On the initial page, the user could decide for either German or English as the language to proceed. Next, the introductory page stated the purpose of the study, the opportunities to share the results with each participant, and the approximate survey length (10-12 minutes). Before starting the survey, all participants agreed to the terms and conditions in which I ensured all participants' confidentiality, anonymity, with respect to the guidelines of the German Research Foundation[6], and voluntariness. Further, the survey showed an information page, which explained that only teams with at least three participants per team can be considered in this study. Next, I introduced a filter question to ensure that the participant had worked with agile methods or practices. Participants who denied that question were directly guided to an ending page. This page explained that the study only considered teams who had experience in using agile methods or practices. Participants who affirmed the question continued regularly with the survey. Further, I assessed the shared mental models, team empowerment, and team reflexivity. Then, participants were asked to answer items regarding team performance, team satisfaction, and team agility. According to Döring and Bortz (2016), demographic questions followed after the last content block of questions. On the next page, participants were asked if they want to receive the results. They could choose between a management summary, attending a webinar, both, and none. To receive the management summary or to attend the webinar, participants had to provide their email addresses. Lastly, the participants received their team- based created link, were reminded to distribute it within the team, and were thanked for their participation. The average duration was 20.63 (SD = 16.45) minutes.

[2]Survey software for students & universities developed by the German company Questback GmbH.

[3]A web-based social network for maintaining existing business contacts and making new business connections.

[4]Either a management summary or to participate in a webinar in which the research process will be explained.

[5]Xing is a German social network in which the members primarily manage their professional contacts.

[6]The German Research Foundation is a registered association that functions as a self-governing institution for the promotion of science and research in the Federal Republic of Germany.

3.3. Measures[7]

All measures of this study were captured with a 5-point-Likert scale ranging from 1 = strongly disagree to 5 = strongly agree, unless stated differently. The bilingual availability of the questionnaire led to a back-translation of all items to German based on Brislin's (1970) suggestions.

3.3.1. Shared Mental Models

The measurement of SMM was distinguished into task-related and team-related. Task- related SMM were assessed by a four-items scale established by Ellwart and Konradt (2007). This scale consists of items to determine the degree of convergence in knowledge structures regarding the performed tasks (one sample item is "I know how the tasks of my team members are related to each other."; $\alpha = .74$). Team-related SMM were assessed with a four items scale established by Konradt et al. (2015). The scale was developed to capture the degree of convergence in knowledge structures regarding the team(work) (one sample item is "I know which team members have expertise in specific areas"; $\alpha = .84$). According to DeChurch and Mesmer-Magnus (2010), I also captured SMM as one construct by using an eight-item scale consisting of the items from both scales. The reported overall Cronbach's alpha was $\alpha = .85$.

3.3.2. Team Empowerment

I measured team empowerment with a 12-item scale from Kirkman et al. (2004) divided into four dimensions with three items in each dimension. One example for meaning of the work of team is "My team feels that its tasks are worthwhile.", for self-determination within a team "My team determines as a team how things are done in the team.", impact of the team "My team performs tasks that matter to this company." and competence of the team "My team can get a lot done when it works hard.". Seibert et al. (2011) and Spreitzer (1995) have shown that those four subdimensions provide a holistic view on whether employees feel psychologically empowered. The four subdimensions were aggregated because they "combine additively to create an overall construct of psychological empowerment. In other words, the lack of any single dimension will deflate, though not completely eliminate, the overall degree of felt empowerment" (Spreitzer, 1995, p. 1444). The reported Cronbach's alpha was $\alpha = .83$.

3.3.3. Team Reflexivity

I measured reflexivity by using a scale introduced by De Jong & Elfring, 2010. The scale consisted of five items (one example regarding work effectiveness is "In our team, we regularly discuss whether we are working effectively together."). The reported Cronbach's alpha was $\alpha = .81$.

[7] The data of the present study were collected as part of a larger project, but I will describe only the variables used in the present research study.

3.3.4. Team Performance

I measured team performance by using the four items scale developed by Lewis (2004). The scale consisted of items regarding the quality of the team output (e.g. "My team's deliverables are of excellent quality.") and the ability of time management (e.g. "My team meets important deadlines on time."). The reported Cronbach's alpha was $\alpha = .73$.

3.3.5. Team Satisfaction

I assessed team satisfaction based on the satisfaction scale of Van Der Vegt, Emans, and Van De Vliert (2000) consisting of two items (one example is "I am satisfied with working in this team"). The reported Cronbach's alpha was $\alpha = .89$.

3.3.6. Team Agility

I assessed the team agility with a self-developed scale based on six agile management practices proposed by Conforto et al. (2014). These agile management practices were classified as fundamentally different from the management practices used in TPM. Specifically, I formulated six statements reflecting each of the six typical agile management practices that are shown in Table 1. The reported Cronbach's alpha was $\alpha = .69$.

3.3.7. Demographic Variables

The measured demographic variables were gender, age, nationality, job position, department, industry, number of employees, and revenue. I measured Gender by using three answer options: 1 = female, 2 = male, 3 = prefer not to answer. In this study, I assessed age with a free text field to enter numbers ("Please enter your age.") and nationality also via a free text field ("Please enter your nationality."). Job position ("What is your current position at your company?") and department ("In which department are you working?") were captured by using a free text field while department was provided with an additional note "Please aggregate your department on the following level e.g., Procurement, Sales, Top Management, R&D etc.". I assessed the item industry by providing a list of 15 industries to choose from plus the option "Other Industry:" with an additional text field. The list of industries can be found in the Appendix F. The number of employees were measured by requesting "Please select the range of number of employees of your company." with the following response options: 1 = 1-49; 2 = 50-249; 3 = 250-499; 4 = 500-999; 5 = 1,000-4,999; 6 = 5,000-9,999 and 7 = > 10,000. The revenue was assessed by asking "Please select the range of number of employees of your company." with the note "Revenue is displayed in million euros." and using the following ranging options: 1 = 0-9; 2 = 10-49; 3 = 50-249; 4 = 250-499; 5 = 500-999; 6 = 1,000-4,999; 7 = 5,000-9,999 and 8 = > 10,000 (see Appendix A).

3.4. Data Analysis

Before I aggregated the data of the team level constructs (SMM, team empowerment, team reflexivity, team agility,

Table 1: Overview team agility construct based on Conforto et al. (2014).

	Item
1.	My team uses the "product vision" concept (e.g., north star).
2.	My team uses simple project plan communication tools and processes.
3.	My team uses an iterative planning approach.
4.	My team develops self-managed and self-directed activities in the project plan.
5.	My team uses self-managed and self-directed plan monitoring and updating activities in the project.
6.	My team frequently applies project plan monitoring and updating processes.

team performance and team satisfaction), I computed several indices to measure within and between-team agreement including r_{wg}[8] and ICC[9] (Bliese, 2000; Klein & Kozlowski, 2000b). According to the literature for team analysis (e.g. Mathieu et al., 2008), aggregating single answers from each team member to the mean value per team is the most common and recommended method. In this study, I followed the r_{wg} definition as "the degree of reliability associated with a single assessment of the group mean" (Bliese, 2000, p. 355). Thus, the r_{wg} indicates whether team members show a similar evaluation of an item varying from 0 to 1 (complete agreement to complete disagreement). All values are displayed in Table 2. All values exceeded cutoff-value of 0.70 (introduced by Lance, Butts, and Michels (2006) as a recommendation value. The ICC values show if a construct has enough consistency within a team so that it can be aggregated to the team level (Bliese, 2000). In this paper, I followed Bliese (2000) definition of ICC: "the degree of reliability associated with a single assessment of the group mean" (Bliese, 2000, p. 355). All variables showed acceptable levels of agreement (A. Cohen, Doveh, & Eick, 2001; LeBreton & Senter, 2008).

Prior to hypothesis testing, I controlled whether the conditions for a linear regression were given (e.g., OLS, multicollinearity, homoskedasticity, etc.). I then tested the hypothesis 1a-4b by using linear regression. For moderation and mediation analysis (Hypothesis 5a – 7d), I used the Process Macro (Hayes, 2013). All the analyses were conducted using SPSS (IBM SPSS Statistics Version 25). Means, standard deviations, and intercorrelations between the study variables are presented in Table 3.

4. Results

4.1. Hypothesis Testing

With respect to the positive relationship between team agility and SMM (Hypothesis 1a), findings did support Hypothesis 1a. In line with my expectations, team agility explained a significant amount of the variance in SMM (F(1,21) = 5.25, R^2 = .20, p = .032). As expected, I also found support for Hypothesis 1b. Team agility explained a significant amount of the variance in team empowerment (F(1,21) = 28.66, p = .000, R^2 = .58).

With respect to the positive relationship between SMM and team empowerment, findings supported Hypothesis 2. In line with my expectations, SMM is pistivels related to team empowerment (F(1,21) = 7.28, R^2 = .26, p = .013).

With respect to the positive relationship between SMM and team performance, findings did support Hypothesis 3a. In line with my expectations, SMM explained a significant amount of the variance in team performance (F(1,21) = 4.45, R^2 = .18, p = .047). As expected, I also found support for Hypothesis 3b, i.e., team empowerment is positively related to team performance (F(1,21) = 16.23, R^2 = .44, p = .001).

With respect to the positive relationship between SMM and team empowerment, findings supported Hypothesis 4a. In line with my expectations, SMM explained a significant amount of the variance in team satisfaction (F(1,21) = 16.23, R^2 = .44, p = .001). As expected, I also found support for Hypothesis 4b, i.e., team empowerment is positively related to team satisfaction (F(1,21) = 6.91, R^2 = .25, p = .016). Tables 4 – 6 show the results of Hypothesis 1a to 4b.

Regarding the positive indirect effect of team agility on team performance through SMM, the results did not support Hypothesis 5a. Contrary to my expectations, the results indicated no significant indirect effect of team agility on team performance via SMM (β = 0.06; 95%, CI [-0.06, 0.39]). With respect to the positive indirect effect of team agility on team satisfaction through SMM, findings did not support Hypothesis 5b. In contrast to my expectations, SMM did not mediate the relationship between team agility and team satisfaction (β = 0.16; 95% CI [-0.03, 0.63]).

Regarding the positive indirect effect of team agility on team performance through team empowerment, the results did not support Hypothesis 6a. Contrary to my expectations, the results indicated no significant indirect effect of team agility on team performance via team empowerment (β = 0.23; 95% CI [-0.10, 0.60]). With respect to the positive indirect effect of team agility on team satisfaction through team empowerment, findings did not support Hypothesis 6b. In contrast to my expectations, team empowerment did not mediate the relationship between team agility and team satisfaction (β = 0.36; 95% CI [-0.19, 0.94]).

Contrary to my expectations, the results indicated no significant moderating effect of team reflexivity on the relationship between SMM and team performance (β = 0.06, p = .871). Thus, I did not find support for Hypothesis 7a. With

[8] r_{wg} = interrater agreement index.
[9] ICC = interclass correlation coefficients.

Table 2: Within-group agreement and between group variance of Study Variables.

Variable	ICC(1)	ICC(2)	F	p	M	rWG(J) Median	rWG(J) N(Teams) <.70
Team Agility	.31	.61	2.58	.002	.80	.91	5
SMM	.19	.44	1.79	.040	.94	.97	3
Team Empowerment	.10	.28	1.39	.161	.96	.96	0
Team Performance	.02	.08	1.08	.391	.85	.91	4
Team Satisfaction	.10	.29	1.40	.153	.77	.91	3
Team Reflexivity	.28	.58	2.37	.005	.83	.88	5

Note. N = 23. ICC = interclass correlation coefficients. rWG(J) = interrater agreement index.

Table 3: Descriptive statistics, reliabilities, and correlations among the study variables.

	Variable	N	M	SD	1	2	3	4	5	6	7	8	9	10
1.	Gender	23	1.61	0.24	NA									
2.	Age	23	30.67	4.76	-.28	NA								
3.	Employees	23	3.91	2.37	.02	.32	NA							
4.	Revenue	23	4.14	2.80	-.05	.27	.96**	NA						
5.	Team Agility	23	3.66	0.49	-.16	.01	-.29	-.27	(.70)					
6.	SMM	23	4.20	0.36	-.09	-.38	-.54**	-.53**	.45*	(.85)				
7.	Team Empowerment	23	4.11	0.31	.06	-.21	-.15	-.09	.76**	.51*	(.83)			
8.	Team Reflexivity	23	3.61	0.58	.01	-.42*	-.25	-.20	.39	.72**	.56**	(.81)		
9.	Team Performance	23	3.89	0.38	.02	-.19	-.19	-.17	.70**	.42*	.66*	.53*	(.73)	
10.	Team Satisfaction	23	4.38	0.49	-.27	-.38	-.17	-.11	.39	.47*	.50**	.69**	.47*	(.89)

Note. Cronbach's α (individual level) are on the diagonal in parentheses. NA = Not Applicable. response categories for gender were: 1 = male, 2 = female and 3 = diverse. Age is displayed in years. Employees and Revenue are displayed in accordance with Appendix A. * p < .05, ** p < .01.

Table 4: Results of Simple Linear Regression Analysis for Hypotheses 1a – 1b.

	SMM			Team Empowerment		
Variable	B	SE B	β	B	SE B	β
Team Agility	0.33	0.14	0.47	1.19	0.22	0.76
R^2		0.20*			0.58***	
F for change in R^2		5.25			28.66	

Note. N = 23. B = unstandardized coefficient. SE B = standard error. β = standardized coefficient. * p < .05. ** p < .01. *** p < .001.

Table 5: Results of Simple Linear Regression Analysis for Hypotheses 3a – 3b.

	Model 1			Model 2		
Variable	B	SE B	β	B	SE B	β
SMM	0.44	0.21	0.42			0.44
Team Empowerment				0.55	0.14	
R^2		0.18*			0.44**	
F for change in R^2		4.45			16.23	

Note. N = 23. B = unstandardized coefficient. SE B = standard error. β = standardized coefficient. * p < .05. ** p < .01. *** p < .001.

respect to team reflexivity moderating the relationship between team empowerment and team performance, the results were not significant (β = 0.32, p = .344). Thus, I also did not find support for Hypothesis 7b.

Table 6: Results of Simple Linear Regression Analysis for Hypotheses 4a – 4b.

	Model 1			Model 2		
Variable	B	SE B	β	B	SE B	β
SMM	0.64	0.26	0.47			0.64
Team Empowerment				0.32	0.12	
R^2		0.22*			0.25*	
F for change in R^2		5.23			6.91	

Note. N = 23. B = unstandardized coefficient. SE B = standard error. β = standardized coefficient. * p < .05. ** p < .01. *** p < .001.

Contrary to my expectations, the results indicated no significant moderating effect of team reflexivity on the relationship between SMM and team satisfaction (β = -0.28, p = .515). Thus, I did not find support for Hypothesis 7c. With respect to team reflexivity moderating the relationship between team empowerment and team satisfaction, the results were not significant (β = 0.08, p = .860). Thus, I did not find support for Hypothesis 7d. Tables 7 – 8 show the results of the mediation analysis to test Hypotheses 5a to 6b.

4.2. Exploratory Analysis

In accordance with the theoretical framework of SMM (Mohammed et al., 2010) and the results of various scholars (Cooke et al., 2003; Lim & Klein, 2006), I further explored all the assumed relationships involving SMM by separating SMM into task-related SMM and team-related SMM. As previous findings have shown, testing for potential dependencies between SMM types can unmask important findings (Smith-Jentsch, Mathieu, & Kraiger, 2005).

With respect to the positive relationship between team agility and task-related SMM, team agility explained a significant amount of the variance in task-related SMM (F(1,21) = 7.43, R^2 = .26, p = .013). However, results did not show support for the positive relationship between team agility and team-related SMM (F(1,21) = 2.61, R^2 = .09, p = .156).

With respect to the positive relationship between task-related SMM and team performance, task-related SMM explained a significant amount of variance in team performance (F(1,21) = 8.85, R^2 = .30, p = .007). However, results did not show support for the positive relationship between team-related SMM and team performance (F(1,21) = 1.23, R^2 = .06, p = .283).

With respect to the positive relationship between task-related SMM and team satisfaction, task-related SMM explained a significant amount of variance in team satisfaction (F(1,21) = 6.07, R^2 = .22, p = .023). However, results did not show support for the positive relationship between team-related SMM and team satisfaction (F(1,21) = 3.31, R^2 = .14, p = .083,).

With respect to the positive relationship between task-related SMM and team empowerment, task-related SMM explained a significant amount of the variance in team empowerment (F(1,21) = 10.10, R^2 = .23, p = .005). However, results did not show support the positive relationship between team-related SMM and team empowerment (F(1,21) = 3.03, R^2 = .12, p = .096).

Regarding the positive indirect effect of team agility on team performance through task- related SMM, the results indicated no significant indirect effect (β = 0.06; 95%, CI [-0.06, 0.39]). With respect to the positive indirect effect of team agility on team performance through team-related SMM, findings did indicate that team-related SMM mediated the relationship (β = 0.01; 95% CI [-0.05, 0.21]).

Regarding the positive indirect effect of team agility on team satisfaction through task- related SMM, the results indicated no significant indirect effect (β = 0.19; 95% CI [-0.08, 0.51]). With respect to the positive indirect effect of team agility on team satisfaction through team-related SMM, findings did indicate that team-related SMM mediated the relationship (β = 0.09; 95% CI [-0.04, 0.47]).

The results indicated no significant moderating effect of team reflexivity on the relationship between task-related SMM and team performance (β = -0.10, p = .801). With respect to team reflexivity moderating the relationship between team-related SMM and team performance, the results were not significant (β = 0.21, p = .489). The results also indicated no significant moderating effect of team reflexivity on the relationship between task-related SMM and team satisfaction (β = -0.47, p = .272). With respect to team reflexivity moderating the relationship between team-related SMM and team satisfaction, the results were not significant (β = -0.06, p = .855)

Given that the small sample team size may be the reason for the non-significant mediation findings (Hypothesis 5a – 6b), I conducted an exploratory investigation into the same mediating relationships on the individual-level (N = 98).

The results showed a positive indirect effect of team agility on team performance through SMM. The results indicated a partial mediation via SMM (β = 0.10; 95% CI [0.03, 0.19]). With respect to the positive indirect effect of team agility on team satisfaction mediated by SMM, the results indicated a partial mediation by SMM (β = 0.13; 95% CI [0.04, 0.26]).

The results showed a positive indirect effect of team agility on team performance via team empowerment, indicating a partial mediation (β = 0.21; 95% CI [0.08, 0.36]). With respect to the positive indirect effect of team agility

Table 7: Results of Mediation Analysis for Hypotheses 5a – 5b (Mediator SMM).

Outcome	Effect of IV on mediator (a)	Unique effect of mediator (b)	Direct effect of IV (c)	Standardized indirect effect	BC 95% CI LL	BC 95% CI UL	Total model effect (a x b + c)	BC 95% CI LL	BC 95% CI UL
Team Performance	0.33*	0.14	0.49**	0.06	-0.06	0.39	0.54***	0.29	0.79
Team Satisfaction	0.33*	0.50	0.22	0.16	-0.03	0.63	0.39	-0.03	0.81

Note. N = 23. IV = independent variable. DV = dependent variable. BC = bias-corrected. LL = lower limit. UL = upper limit. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 8: Results of Mediation Analysis for Hypotheses 6a – 6b (Mediator Team Empowerment).

Outcome	Effect of IV on mediator (a)	Unique effect of mediator (b)	Direct effect of IV (c)	Standardized indirect effect	BC 95% CI LL	BC 95% CI UL	Total model effect (a x b + c)	BC 95% CI LL	BC 95% CI UL
Team Performance	0.49***	0.37	0.36	0.23	-0.10	0.60	0.54***	0.29	0.79
Team Satisfaction	0.49***	0.74	0.03	0.36	-0.19	0.94	0.39	-0.03	0.81

Note. N = 23. IV = independent variable. DV = dependent variable. BC = bias-corrected. LL = lower limit. UL = upper limit. * $p < .05$; ** $p < .01$; *** $p < .001$.

on team satisfaction mediated by team empowerment, the results indicated a partial mediation by team empowerment ($\beta = 0.28$; 95% CI [0.14, 0.44]). Tables 9 – 10 show the results of the exploratory mediation analysis.

5. Discussion

The application of team agility has developed into a mainstream approach in the workplace context in order to cope better with change (Denning, 2018; Tessem, 2014). Although a principle of the Agile Manifesto holds that individuals and interaction take precedence over tools and processes (Beck et al., 2001), we still lack insights into how the human-side functions. Moreover, one critical issue is the absence of a theoretical basis of team agility that puts it into the context of teamwork research. Therefore, the main goal of the current study was to examine the impact of team agility on team effectiveness using SMM and team empowerment as potential mediators in the relationship between team agility and team outcomes. Further, using team reflexivity as a moderator in each relationship between mediators and team outcomes, my work provides an important step towards a new theoretical viewpoint of team agility. In order to theorize and empirically investigate the impact of team agility, I draw on the IMO-framework (Ilgen et al., 2005) and examined agile teams in various industries without limiting my study to a specific agile method. Furthermore, this study builds on specific management practices used in team agility (Conforto et al., 2016).

The results indicated a positive relationship between team agility and team effectiveness. These findings are in line with previous research (Schmidt et al., 2014), but also extend it with a different study focus. The present study focused more on the investigation of team agility in general than on limited specific agile methods. Results also showed that team agility worked as an antecedent of both, SMM and team empowerment, that facilitate team effectiveness. Besides, results showed a positive relationship between these two emergent states. In line with previous research (Kirkman et al., 2004; Lim & Klein, 2006; Seibert et al., 2011), my findings showed a positive relationship between all four relationships of the mediators (SMM and team empowerment) and the outcomes (team performance and team satisfaction). Further, the results showed no mediation effects for SMM and team empowerment in the relationship to team performance and team satisfaction. However, the results have indicated that a total model effect exists for the relationship between team agility and team performance mediated by SMM and team empowerment. No moderating effect of team reflexivity on the relationship between SMM and team performance respectively team satisfaction was supported by the results. Also, the findings did not support a moderating effect of team reflexivity on the relationship between team empowerment and team performance respectively team satisfaction.

5.1. Interpreting the Results

5.1.1. Team Agility and SMM

The current study revealed that team agility positively impacted SMM. These findings validate the conceptual analysis of Yu and Petter (2014): team agility is beneficial to the development of SMM. Further, it extends the research of Schmidt et al. (2014) regarding the impact of team agility on SMM. They focused on software development teams that applied specific agile practices or methods. Thus, my results imply that the building of SMM through team agility is not bound to certain industries (e.g., software development) or methods (e.g., Scrum). These results can be explained and understood in two ways. First, the results are in line with the argumentation that team agility impacts the building of SMM because it encapsulates important antecedents

Table 9: Results of Exploratory Mediation Analysis for (Mediator SMM).

Outcome	Effect of IV on mediator (a)	Unique effect of mediator (b)	Direct effect of IV (c)	Standardized indirect effect	BC 95% CI LL	BC 95% CI UL	Total model effect (a x b + c)	BC 95% CI LL	BC 95% CI UL
Team Performance	0.29***	0.34*	0.41***	0.10	0.03	0.19	0.51***	0.33	0.68
Team Satisfaction	0.29***	0.53**	0.35*	0.13	0.04	0.26	0.50***	0.28	0.71

Note. N = 98. IV = independent variable. DV = dependent variable. BC = bias-corrected. LL = lower limit. UL = upper limit. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 10: Results of Exploratory Mediation Analysis for (Mediator team empowerment).

Outcome	Effect of IV on mediator (a)	Unique effect of mediator (b)	Direct effect of IV (c)	Standardized indirect effect	BC 95% CI LL	BC 95% CI UL	Total model effect (a x b + c)	BC 95% CI LL	BC 95% CI UL
Team Performance	0.46***	0.46**	0.30**	0.21	0.08	0.36	0.51***	0.33	0.68
Team Satisfaction	0.46***	0.72***	0.17	0.28	0.14	0.44	0.50***	0.28	0.71

Note. N = 98. IV = independent variable. DV = dependent variable. BC = bias-corrected. LL = lower limit. UL = upper limit. * $p < .05$; ** $p < .01$; *** $p < .001$.

of SMM. In detail, the iterative planning approach of team agility can work as one factor to facilitate SMM as Orasanu (1994) has shown that (effective) planning facilitates the building of SMM. In fact, regular planning is also associated with frequent communication and alignment what can lead to a better coordination. That means the iterative planning approach of team agility contributes to a shared understanding of the next steps and ensures that the team is on the same page. Second, team agility can be interpreted as a form of structured self-reliance. A team agrees on intra-team related working guidelines (e.g., Scrum framework) and adopts them. Adopting these kinds of working norms is comparable to committing to one philosophy. The more pronounced this takes place, the greater the resulting shared philosophy on the working model. Therefore, a shared philosophy can foster clarity regarding teamwork (e.g., how to work, how to interact). The common philosophy ensures one mindset among team members, which can contribute positively to the building of SMM (Klimoski & Mohammed, 1994). Moreover, team agility can be understood as an important antecedent of SMM.

5.1.2. Team Agility and Team Empowerment

Another important finding was that team agility positively impacted team empowerment. These findings extend previous research studies (Tessem, 2014), which showed greater individual empowerment through an increased flow of information in agile work teams compared to traditional, more documentation-oriented work teams. My study adds an additional layer of inquiry by examining the issue at the team level. My findings can be interpreted in three ways. First, the results are in line with the argumentation that team agility impacts team empowerment because it encapsulates important antecedents of team empowerment. According to Seibert et al. (2011), socio-political support leads to greater team empowerment. Therefore, it seems like teams with a higher degree of freedom (respectively team agility) experience a higher feeling of competence because they feel socio-political support. Socio-political support can be expressed by providing employees with resources that are either material, social, or psychological (Spreitzer, 1996). Thus, one explanation could be that team agility provides some kind of socio-political support in form of psychological resources, so that it positively impacts team empowerment.

Second, it is possible that advanced team agility results in a greater feeling of team empowerment through participative decision-making and shared leadership. One important component of team agility is the ability to distribute and prioritize tasks together in the team without being told by management (Moe et al., 2008). Therefore, it can enhance the team's feeling of responsibility and ownership. These feelings perhaps reflect an increased experience of autonomy and competence. This working approach can also be characterized as one form of high-management performance practices. High-management performance practices work as an antecedent of team empowerment. They are intended to provide employees with better control over their work (Seibert et al., 2011).

Third, it is possible that team agility advocates high information sharing, which fosters the team's feeling of empowerment. Teams who work agile aim to provide each team member with as much information as possible, often daily (e.g., daily stand-ups). Another way of information sharing that they use are so called information radiators, such as burn charts (Williams & Cockburn, 2003). Information radiators are used to easily make information on the status of the project visible and transparent to all. In addition, the increased exchange of information eventually has a positive effect on the team's feeling of competence, as it has a better basis for making decisions.

Overall, these findings may help us to understand the positive relationship of team agility and team empowerment. My study contributes to a better theoretical understanding of the impact and consequences of team agility. Further, the findings relate team agility to the literature of team effectiveness by showing team agility as a possible antecedent of team empowerment.

5.1.3. SMM and Team Empowerment

The present study showed a positive relationship between the cognitive emergent state SMM and motivational emergent state team empowerment. My results can be explained in the following ways: High SMM similarity explains that the team is consistent in how to approach a task and how to interact as a team. Having this knowledge possibly fosters the collective belief that they have the competence to master the upcoming tasks. Further, it improves the internal knowledge basis of the team. A team gains knowledge regarding the responsibilities, competencies, and feelings of each team member. In this way, it can contribute to a team's greater empowerment. Using the internal knowledge can support the collective establishment of objectives and the creation of a supportive work environment. In this way, it can contribute to a team's greater feeling of self-determination. Fulfilling the aim of self-determination is an important component to empower a team (Seibert et al., 2011). Overall, these findings can enhance our knowledge about the relationship of cognitive and motivational states. There is an interaction between these two emergent states.

5.1.4. Mediational Role of SMM

My assumptions that SMM positively mediate the relationship of team agility and team outcomes were not supported. These findings could be explained by the following factors: First, team agility and SMM work too similar to identify a mediational effect. Team agility works as a kind of philosophy with a form of structured self-reliance. That means teams commit to a special way of working, which allows them to have a certain degree of autonomy. Therefore, it could be that team agility requires SMM and does not foster it. Having a similar understanding of how to operate as a team is a necessary prerequisite for team agility. Because the responsibilities are shared from the start, the team members must be able to rely on each other (Moe et al., 2008). If a team member is unable to predict the next steps which other team members are planning and how they can be supported to be successful, it may diminish the applicability of team agility. Further, it is possible that team agility enables increased communication, coordination, and visibility of tasks (Dingsøyr, Dybå, & Moe, 2010; Williams & Cockburn, 2003) that may render the effects of SMM obsolete. While SMM are developed that team members can better predict and expect what other team members are doing, the nature of cooperation in agile working teams might substitute that through the open and free communication.

Second, it is also possible that different emergent states might explain the relationship between team agility and team outcomes. For example, team trust could be investigated as an additional mediator. Team trust refers to the willingness of each team member to believe that the others will perform their tasks as expected without controlling them (Mayer et al., 1995). Team agility relies on self-organizing and collaboration in the team. Self-organizing, agile teams must be able to rely on each other because each team member is responsible for his or her own tasks. When one team member is not delivering the results on time, it will negatively affect the whole team (McHugh, Conboy, & Lang, 2011). Further, team agility promotes open communication and problem sharing that is needed to discuss mutually next steps and to reflect on the past performance period (Moe et al., 2010). When teams lack trust, it has a negative impact on their collaboration. Team members will shift their focus to monitoring and controlling each other rather than sharing ideas (Salas, Sims, & Burke, 2005). Thus, team trust is an emergent state that could be better suited to impact the relationship between team agility and team outcomes.

Moreover, my exploratory analysis supported that SMM positively mediates the relationship of team agility and team performance. It is possible that the constructs were perceived differently on an individual-level than on a team-level so that their relationship worked better on an individual-level. One reason for this could be that team agility impacts some individuals more than others. Those promote their development of SMM to greater extent compared to other team members. This effect could be neglected on a team-level if others feel the opposite or did not experience the effect as strongly. This could be also related to their individual knowledge and experience with team agility. Team members who have already worked agile over a longer period could not perceive that effect because they already developed a mental model which they also assume the others to have.

Overall, the results demonstrate that SMM did not mediate the relationship of team agility team performance or team satisfaction. Thus, the need to test further variables which may mediate these relationships persists. In agreement with LePine et al. (2008), other team processes should be examined.

5.1.5. Mediational Role of Team Empowerment

Contrary to my expectations, team empowerment did not mediate the positive relationship of team agility and team outcomes. These findings could be explained by the following factors: First, in this relationship, team agility already requires team empowerment. Teams that are self-managed and self-organized need to be empowered in the first place. When a team does not believe in its competence and self-determination, team agility would not be the right working approach for them. In addition, team members need to have a strong will to overtake responsibility if they want to work successfully. In other words, some components of team empowerment could be an important antecedent of the willingness to work agile. Another explanation could be that team agility enables feelings of self-determination and competence through the self-managing nature so that team empower-

ment is not that essential to reach high outcomes. Thus, it can be that team agility per se captures some effects of team empowerment in such a way that it diminishes the relationship via team empowerment to team outcomes.

Second, it is also possible that different emergent states might explain the relationship between team agility and team outcomes. For example, team cohesion could be investigated as an additional emergent state. Team cohesion refers to the commitment of the entire team to hold together and share agreement on team responsibilities and team tasks (Mullen & Copper, 1994). It implies that the team is aligned considering their priorities. Frequent communication supports the integration of all team members into a team and create a cohesive atmosphere (McHugh et al., 2011). Information sharing through information radiators increases the awareness of the entire project team regarding what is the status quo (Williams & Cockburn, 2003). Through reducing the lack of awareness, team cohesion should be impacted by team agility (Whitworth & Biddle, 2007). Thus, team cohesion is an emergent state that could be better suited to impact the relationship between team agility and team outcomes.

In addition, the exploratory analysis showed that team empowerment positively mediates the relationship of team agility and team performance. It is possible that the constructs were perceived differently on an individual-level compared to the team-level so that their relationship worked on an individual-level. One reason for this could be that the individuals differ in their perception of empowerment. Whereas one could perceive the self-organizing nature of team agility as self-determination, the other could perceive that more as a burden. Further, their feelings may differ along the team empowerment dimension of meaningfulness. For one individual, it could be more related to the way of working whereas someone else could relate it more to the topic they work on. That means the dimensions of team empowerment can differ among individuals such that the effect of team empowerment was not observable on a team-level. This could also be related to their individual knowledge and experience with team agility. Team members who already have experiences with team agility may have a different understanding of team empowerment compared to others.

Overall, the results demonstrate that team empowerment did not mediate the relationship of team agility, team performance, or team satisfaction. Thus, the need to test further variables which may mediate these relationships persists. In agreement with LePine et al. (2008), other team processes should be examined.

5.1.6. Moderating Role of Team Reflexivity

This study did not find any moderating effect of team reflexivity in the relationship between the mediators and the team outcomes. That could be explained in three ways. First, I argued that team reflexivity creates conditions in which teams embrace a higher level of self- awareness to be better aligned in terms of team interactions and tasks. But this may hold only for teams that work traditionally and not agile. Two core components of team agility is frequent communication and regular planning (Dybå & Dingsøyr, 2008). These components can counter the effect of team reflexivity to create awareness and an open atmosphere of sharing experiences because those conditions are present the whole time. In an agile work environment, teams constantly challenge and adapt their solution approach (Schwaber, 2004).

Second, agile working teams are may limited in their functional dimension, thus, a lack of diversity can affect the impact of team reflexivity (Yang et al., 2020). Poor diversity in a team can hinder a broad range of information because the degree of cognitive overlap increases, and teams miss necessary perspectives. Thus, team reflexivity cannot unfold its effects and contribute to a work environment in which teams diversely assess their status quo.

Third, teams may require some important drivers of effective team reflexivity, like feedback availability, psychological safety, and empowering leadership (Otte et al., 2019). When these antecedents are not met, team reflexivity is likely to fail. Teams may spend too much time on superficial considerations and omit important topics or come to false conclusions (Konradt et al., 2016). Thus, team reflexivity cannot embrace conditions such that it positively impacts the relationship between emergent states and team outcomes. In sum, in teams with an agile working approach, team reflexivity possibly conflicts with the general openness of teams in such a way that it cannot truly unfold its benefits in these conditions. Team agility seems to encompass at least parts of team reflexivity so that it diminishes the impact of team reflexivity. Interestingly, the analysis revealed that reflexivity was negatively related to age, meaning that the older the participants, the less they reflected. A closer examination of this finding offers an interesting avenue for future research, as it could provide important insights into how willingness to adapt to a particular practice changes with age.

5.1.7. Task-related SMM and team-related SMM

According to Mohammed et al. (2010), previous studies have too often disregarded the multidimensional nature of SMM. Therefore, this study examined additionally task-related SMM and team-related SMM separately in an exploratory manner. The results showed different direct effects for both SMM dimensions on all variables. These findings are in line with previous studies showing different effect sizes when separating the SMM dimensions into task-related and team-related (Cooke et al., 2003; Lim & Klein, 2006; Mathieu et al., 2000; Mathieu et al., 2005). A possible explanation is that this is related to the context domain the teams operate in. The idea of SMM is mostly applied and investigated in a more action-based environment, like military teams (Mohammed et al., 2010), where a task-related SMM is may more important, because the interaction and communication follows strict hierarchies and principles. Similar for top-management teams in a corporate world, they would probably focus more on how to operationally conduct a task because the communication is associated to be time consuming and to be responsible for decision delays (Smith et al., 1994). Both team contexts have in common that they

normally face high-complex tasks and where the teams are primarily evaluated by their performance. Thus, teams are rather concentrating on finding solutions by taking all internal and external circumstances into consideration than aligning the way how they work together. I assume that results could differ in work contexts where teams are not always confronted with high pressure situations.

Further, team agility positively impacted task-related SMM, but not team-related SMM. It is possible that task-related SMM is more important in an agile working context. One reason for this could be that the focus of the iterative planning approach is on continuous adaptation to work objectives and performance requirements. The whole team regularly discusses what has been achieved in the past performance period and what has to be next in order to fulfill the project goals. Additionally, agile work teams often use the product vision concept (Highsmith, 2004). The product vision concept is designed for teams to define their overarching project goal. Thus, these measures contribute to a better shared understanding of the tasks. Task-related SMM refer to a team similarity in knowledge structures regarding which task should be performed next and how should it be performed to reach the goal. Therefore, team agility works better as an antecedent for task-related SMM than for team-related SMM. However, the interaction between both forms of SMM lead me to conclude that they combine in some manner to impact team outcomes. Presumably, these effects can change over time and become more aligned for teams working together over longer periods.

5.2. Theoretical Implications

This study makes several important contributions to the existing agility literature and teamwork research, linking the two. First, my study extends previous work on the impact of team agility on team performance (Campanelli & Parreiras, 2015; Schmidt et al., 2014; Serrador & Pinto, 2015) beyond focusing on explicit agile methods (e.g., Scrum) or certain industries (e.g., software development). By introducing a measurement approach to assess team agility, it opens the path to a broader investigation of team agility beyond the software development context. I argue that the application of team agility becomes more and more popular in the organizational context (Denning, 2018; Tessem, 2014) without knowing whether it delivers the expected impact or not. Thereby, it helps to further explore and understand how the human side of team agility functions.

Second, my findings successfully integrate the agile literature and teamwork research by demonstrating team agility as an essential driver for the emergence of team effectiveness. In contrast to most previous literature (Dingsøyr et al., 2010; Dingsøyr et al., 2016; Laanti, 2013; Moe et al., 2010), I provide empirical support for the theorized relationship. Besides, the implementation of team agility in the IMO-framework provides a strong theoretical basis for explaining the impact of team agility at the team level. In this way, my study contributes to prior literature on team effectiveness by showing that SMM and team empowerment are related to team agility. Moreover, I respond to the calls of Stray et al. (2018), who demanded, from the agile literature stream, a more detailed investigation of mechanisms that enable agile work teams who coordinate their actions and behaviors as a response to internal and external circumstances. Further, Mathieu et al. (2017) asked, from a teamwork research perspective, for deeper insights regarding teaming in new work arrangements, like the agile work context. By examining team agility with a focus on emergent states as a consequence of it, this study also provides additional insights into how the human side of team agility works (Grass, Backmann, & Hoegl, 2020; Schmidt et al., 2014). By clearly outlining the relationship of team agility and emergent states, this study goes beyond the existing literature. While previous literature relied on conceptualizations (Moe et al., 2008; Yu & Petter, 2014) or a limited investigation perspective (Hoda & Murugesan, 2016; Schmidt et al., 2014), this study could validate that relationship on a broader scale. Thus, my results can form the basis for a more detailed analysis of team agility by providing more theory-based ideas of how team agility functions in organizational contexts. Third, this study helps us to better understand and critically review the concept of team reflexivity in the context of team agility. Team reflexivity is a well-known and investigated concept that is positively related to team outcomes (Konradt et al., 2016). As team reflexivity is anchored in the idea of team agility (Beck et al., 2001; Moe et al., 2010), I argued that it should be investigated if it strengthens the relationship of emergent states (SSM and team empowerment) and outcomes in an agile work context. Surprisingly, in my study, team reflexivity did show any effect. One of the issues that emerges from these findings is if the impact of team reflexivity is reduced in an agile work environment. One reason could be that an agile work environment fosters an open communication and discussion forum on a regular basis. If so, there is a necessity to reevaluate whether the concept of team reflexivity holds up in these work environments. This study did not distinguish between self-regulated reflexivity or guided reflexivity. They differ in their initiation. While self-regulated reflexivity focuses on a dynamic process that happens by chance, guided reflexivity focuses on a structured and pre- arranged teamwork session (Otte et al., 2019). Team agility relies on a more structured reoccurring team reflexivity process. In combination with the open-minded teamwork policy that should foster self-regulated reflexivity, it can be assumed that the positive effect of the specific team reflexivity session decreases. Further, time is an important factor, as teams need to develop over time to implement a meaningful reflexivity process.

Lastly, this study also contributes to SMM research by enhancing our knowledge of the relation to team empowerment and by further supporting our understanding of task-related and team-related SMM. By doing so, I respond to the call of Mohammed et al. (2010), who suggested to further extend the criteria and context of empirical research of SMM and to measure multiple SMM content domains. Results extend literature by showing a relationship between SMM and team empowerment, demonstrating the impact on team satisfac-

tion. However, the causal effects remain unclear. In the context of decision-making management teams, the task-related SMM is the predominant component that positively impacts team outcomes, such as team performance or team satisfaction. Knowing that, the study helps to better clarify the roles of both mental models.

5.3. Practical Implications

The results of my study also provide valuable insights for various practitioners, including decision-makers, project managers, human resource managers, or, more broadly, organizations. In general, my findings should encourage organizations to adopt team agility as a project management approach. Team agility helps to increase SMM and team empowerment of teams. For example, SMM are especially useful for teams that operate in a complex, dynamic, and uncertain environment (Cannon-Bowers et al., 1993). Teams that share an internal knowledge base and agree on what tasks need to be completed and with whom actions need to be coordinated consequently demonstrate better team performance. Further, organizations benefit from team agility as it impacts team empowerment. Team empowerment is a newly arising important factor for the future workforce. Graduates strive for self- determination, meaningfulness, and making an impact. Implementing team agility as a work approach can support organizations in remaining attractive for future employees as it is a state- of-the-art working approach.

Human resource manager should seek for employees who exhibit skills and competencies that are inherent in agile work teams, such as the ability to self-organize and self-manage. Further, they can design developments paths for employees to practice those abilities. In alignment with the corresponding organization, they can advertise their firm with the benefits of team agility like perceived team empowerment or creating an internal knowledge basis with their colleagues (SMM).

Project managers can make use of the results because they offer them team agility as a new project management approach. Implementing (parts of) team agility supports the development of SMM and the feeling of team empowerment. When they assemble a new team, they can use the team agility approach to bring the team on the same page regarding their task- and teamwork. Moreover, the self-organizing approach can be used to take some pressure of the project manager by distributing responsibilities. However, they should be aware of the fact that team reflexivity is not always the way to go for agile working teams. Moreover, it should be considered how and when teams reflect. Even though agile practitioners propose a routine to reflect (Abrahamsson et al., 2002), my findings should serve as caveats. The beneficial implications of team reflexivity can depend on various impact factors, like task complexity, team development stage or type, or reflexivity (self-regulated vs. guided). Therefore, project managers should experiment with the optimal way of when and how to apply team reflexivity.

The study is beneficial for decision-makers because they can better assess the impact and effectiveness of team agility in a more generalized context (e.g., compared to software development context). Through developing a better understanding of how the mechanisms triggered by team agility work, decision-makers are provided with an idea of how the power of team agility unfolds. Therefore, they should keep in mind that results proved that teams with greater team agility also showed greater effects. Examining the impact of team agility on SMM and team empowerment provides decision makers with valuable insights into what other factors they should consider when evaluating the benefits of team agility.

Overall, this study provides evidence that team agility positively impacts the SMM and team empowerment, that, in turn, leads to higher team effectiveness. The assumption that team agility encompasses great features for improving collaboration within the team could be validated.

5.4. Limitations and Future Research

Despite the promising results, this study does not come without some limitations. However, some of these can also provide opportunities for future research. First, this research is based on a relatively small sample size. A small sample size comes with low statistical power and is susceptible to a type II ("false positive") error. Thus, the generalizability is limited, and the data must be interpreted cautiously. Future research may attempt to replicate this study with a larger and more international sample size to validate my findings. I assume that different cultures tend to stick differently to team agility management practices, which could reveal a more comprehensive view of single effects.

Second, the cross-sectional survey design does not allow to draw conclusions about causality because all variables were assessed simultaneously. But teams go through a process of development (e.g., forming, storming, norming, performing) that leads to a change in their interactional behavior (Tuckman, 1965) and team emergent states change over time (Marks et al., 2001). For example, it is conceivable that relationships differ for new teams compared to teams that have been working together for a while. Future research should investigate these complex relationships in longitudinal studies. Besides, an interesting new research avenue includes studying team agility and the emergence of SMM and team empowerment over time. Next, given that this study relies only on self-reporting measures, it can result in a self-reporting measure bias. Participants might rather give desired than honest answers. In addition, various questions may have been rather difficult to answer accurately to oneself, like evaluating the team agility or team performance. Participants may answer team agility better or worse if they like or dislike this approach. In terms of evaluating their own team performance, participants may be more likely to rate their performance as they would like it to be. In particular, future attempts should be made to measure team performance either by external raters or on the basis of more objective criteria. To reduce the self-reporting bias completely future research should aim to observe actual behavior. One idea would be to use the state space grids method to better cap-

ture and analyze team dynamics (Meinecke, Hemshorn de Sanchez, Lehmann-Willenbrock, & Buengeler, 2019).

The aggregation of individual data on team-level showed that the ICC(2) estimates are not ideal compare to the standard .70 benchmark. But in accordance with Bliese (1998), it is not unregular that ICC(2) values are rather low when teams are small. By choosing three as a minimum amount of team members to form a team, the team average of this study is with 3.48 individuals per team rather small. Bliese (1998) argued that a possible explanation for low mean team reliability can be the underestimation of actual team-level relationships. However, researchers should note that both the ICC(1) and the rWG values are within a common and acceptable range.

Lastly, I used a not established construct to measure team agility in a quantitative empirical research design. Nonetheless, the reliability and validity checks (Cronbach's $\alpha = 0.69$) were rather acceptable and build on the identifications of typical agile project management practices (Conforto et al., 2014). However, future research should strive to further improve the proposed construct by conducting qualitative research (e.g., interview agile practitioners), a confirmatory analysis, and ultimately, validate and introduce the construct. In addition, it should be investigated if team agility could be distinguished into teamwork and taskwork dimensions, like the traditional literature on work teams does (Fisher, 2014). These caveats should be kept in mind when interpreting my results.

Countering these limitations, the study provided some strengths that should be noticed as well. One important strength is the investigation of teams from different industries as well as teams with different functional roles. This allows us to generate a broader perspective of team agility. In total, teams varied across 14 different industries and 13 functional roles. In addition, I was able to assess teams consisting of various team sizes, ranging from minimum three to maximum six people. Investigating different team sizes is important because it can reveal vital information about dynamics of distinct teams.

5.5. Conclusion

More and more organizations are adopting team agility as a working approach to cope with the arising challenges of a rapidly changing work environment and to deal with increasing complexity (Conforto et al., 2014; Denning, 2018). Previous research did not sufficiently investigate the impact of team agility on team effectiveness, and especially how the relationship between team agility and emergent states works. Therefore, the goal of my research study was to combine the new literature stream on team agility with established theories from teamwork research.

My results indicate that team agility has a positive impact on both SMM and team empowerment. Team agility incorporates important antecedents of SMM, like iterative planning, that foster the development of SMM. Moreover, it is the nature of agile teams to work self-organized and self-managed within a certain framework. This form of structured self- reliance can facilitate the teams' perception of empowerment through increased feelings of autonomy and competence. Further, the results showed that the role of team reflexivity in an agile work environment remains unclear. Although team reflexivity is an essential component of team agility, it did not create conditions that fostered a relationship between SMM, respectively team empowerment, and team outcomes. These findings help us to understand and explain how team agility can contribute to team effectiveness.

Overall, my study takes a significant step towards the integration of the literature stream on team agility and teamwork research. My work can provide the foundation for further research and theory on team agility as an approach to increase team effectiveness. The insights into the relationship between team agility and SMM and team empowerment improves our understanding of how the human-side functions in an agile work environment.

References

Abrahamsson, P., Salo, O., Ronkainen, J., & Warsta, J. (2002). *Agile software development methods: Review and analysis* (Tech. Rep.). VTT Technical report. (Vol. 478)

Al Jenaibi, B. (2010). Job satisfaction: Comparisons among diverse public organizations in the UAE. *Management Science and Engineering*, *4*(3), 60–79.

Argote, L., Gruenfeld, D., & Naquin, C. (1999). Group learning in organizations. In M. E. Turner (Ed.), *Groups at work: Advances in theory and research*. Lawrence Erlbaum.

Balkundi, P., & Harrison, D. A. (2006). Ties, leaders, and time in teams: Strong inference about network structure's effects on team viability and performance. *Academy of Management Journal*, *49*(1), 49–68.

Beck, K., Beedle, M., van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., ... Thomas, D. (2001). *Manifesto for agile software development*. http://www.agilemanifesto.org/.

Bliese, P. D. (1998). Group size, ICC values, and group-level correlations: A simulation. *Organizational Research Methods*, *1*(4), 355–373.

Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multileveltheory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 349–381). Jossey-Bass.

Boehm, B., & Turner, R. (2003). *Balancing Agility and Discipline: A Guide for the Perplexed*. Addison-Wesley.

Boehm, B., & Turner, R. (2005). Management challenges to implementing agile processes in traditional development organizations. *IEEE Software*, *22*(5), 30–39.

Bommer, W. H., Johnson, J. L., Rich, G. A., Podsakoff, P. M., & MacKenzie, S. B. (1995). On the interchangeability of objective and subjective measures of employee performance: A meta-analysis. *Personnel Psychology*, *48*(3), 587–605.

Bretz, R. D. J., & Judge, T. A. (1994). Person-organization fit and the theory of work adjustment: Implications for satisfaction, tenure, and career success. *Journal of Vocational Behavior*, *44*(1), 32–54.

Burke, C. S., Stagl, K. C., Klein, C., Goodwin, G. F., Salas, E., & Halpin, S. M. (2006). What type of leadership behaviors are functional in teams? a meta-analysis. *The Leadership Quarterly*, *17*(3), 288–307.

Campanelli, A. S., & Parreiras, F. S. (2015). Agile methods tailoring – a systematic literature review. *Journal of Systems and Software*, *110*, 85–100.

Campion, M. A., Medsker, G. J., & Higgs, A. C. (1993). Relations between work group characteristics and effectiveness: Implications for designing effective work groups. *Personnel Psychology*, *46*(4), 823–847.

Cannon-Bowers, J. A., Salas, E., Blickensderfer, E., & Bowers, C. A. (1998). The impact of cross-training and workload on team functioning: A replication and extension of initial findings. *Human Factors*, *40*(1), 92–101.

Cannon-Bowers, J. A., Salas, E., & Converse, S. A. (1993). Shared mental models in expert team decision making. In N. J. Castellan Jr. (Ed.), *Current Issues in Individual and Group Decision Making* (pp. 221–246).

Chen, G., Kirkman, B. L., Kanfer, R., Allen, D., & Rosen, B. (2007). A multilevel study of leadership, empowerment, and performance in teams. *Journal of Applied Psychology*, *92*(2), 331–346.

Chen, G., Sharma, P. N., Edinger, S. K., Shapiro, D. L., & Farh, J.-L. (2011). Motivating and demotivating forces in teams: Cross-level influences of empowering leadership and relationship conflict. *Journal of Applied Psychology*, *96*(3), 541–557.

Chou, L.-F., Wang, A.-C., Wang, T.-Y., Huang, M.-P., & Cheng, B.-S. (2008). Shared work values and team member effectiveness: The mediation of trustfulness and trustworthiness. *Human Relations*, *61*(12), 1713–1742.

Cohen, A., Doveh, E., & Eick, U. (2001). Statistical properties of the rWG (J) index of agreement. *Psychological Methods*, *6*(3), 297–310.

Cohen, S. G., & Bailey, D. E. (1997). What makes teams work: Group effectiveness research from the shop floor to the executive suite. *Journal of Management*, *23*(3), 239–290.

Combs, J., Liu, Y., Hall, A., & Ketchen, D. (2006). How much do high-performance work practices matter? a meta-analysis of their effects on organizational performance. *Personnel Psychology*, *59*(3), 501–528.

Conboy, K., & Fitzgerald, B. (Eds.). (2004). *Toward a Conceptual Framework of Agile Methods: A Study of Agility in Different Disciplines*. Springer Verlag.

Conforto, E. C., Amaral, D. C., da Silva, S. L., Di Felippo, A., & Kamikawachi, D. S. L. (2016). The agility construct on project management theory. *International Journal of Project Management*, *34*(4), 660–674.

Conforto, E. C., Salum, F., Amaral, D. C., Da Silva, S. L., & De Almeida, L. F. M. (2014). Can agile project management be adopted by industries other than software development? *Project Management Journal*, *45*(3), 21–34.

Cooke, N. J., Kiekel, P. A., Salas, E., Stout, R., Bowers, C., & Cannon-Bowers, J. (2003). Measuring team knowledge: A window to the cognitive underpinnings of team performance. *Group Dynamics: Theory, Research, and Practice*, *7*(3), 179–199.

Cooper, R. G., & Sommer, A. F. (2016). The agile-stage-gate hybrid model: A promising new approach and a new research opportunity. *Journal of Product Innovation Management*, *33*(5), 513–526.

Creswell, J. W. (2008). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (3rd ed.). Pearson.

DeChurch, L. A., & Mesmer-Magnus, J. R. (2010). Measuring shared team mental models: A meta-analysis. *Group Dynamics: Theory, Research, and Practice*, *14*(1), 1–14.

Deci, E. L., & Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. *Journal of Research in Personality*, *19*(2), 109–134.

De Jong, B. A., & Elfring, T. (2010). How does trust affect the performance of ongoing teams? the mediating role of reflexivity, monitoring, and effort. *Academy of Management Journal*, *53*(3), 535–549.

Denning, S. (2013). Why Agile can be a game changer for managing continuous innovation in many industries. *Strategy & Leadership*, *41*(2), 5–11.

Denning, S. (2018). How major corporations are making sense of Agile. *Strategy & Leadership*, *46*(1), 3–9.

DeSimone, J. A., Harms, P. D., & DeSimone, A. J. (2015). Best practice recommendations for data screening. *Journal of Organizational Behavior*, *36*(2), 171–181.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode Surveys* (4th ed.). John Wiley & Sons.

Dingsøyr, T., Dybå, T., & Moe, N. B. E. (2010). *Agile Software Development* (Vol. 50) (No. 9-10). Springer.

Dingsøyr, T., Fægri, T. E., Dybå, T., Haugset, B., & Lindsjørn, Y. (2016). Team performance in software development: Research results versus agile principles. *IEEE Software*, *33*(4), 106–110.

Dingsøyr, T., Falessi, D., & Power, K. (2019). Agile development at scale: The next frontier. *IEEE Software*, *36*(2), 30–38.

Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Springer Berlin Heidelberg.

Dybå, T., & Dingsøyr, T. (2008). Empirical studies of agile software development: A systematic review. *Information and Software Technology*, *50*(9-10), 833–859.

Edwards, B. D., Day, E. A., Arthur Jr, W., & Bell, S. T. (2006). Relationships among team ability composition, team mental models, and team performance. *Journal of Applied Psychology*, *91*(3), 727–736.

Ellwart, T., & Konradt, U. (2007, May). *Explicit and implicit team coordination: Influences of planning and shared mental models on team conflicts and coordination success*. (Paper presented at the 13th Congress of Work and Organizational Psychology, Stockholm)

Ensley, M. D., & Pearce, C. L. (2001). Shared cognition in top management teams: Implications for new venture performance. *Journal of Organizational Behavior*, *22*(2), 145–160.

Fisher, D. M. (2014). Distinguishing between taskwork and teamwork planning in teams: Relations with coordination and interpersonal processes. *Journal of Applied Psychology*, *99*(3), 423–436.

Ghezzi, A., & Cavallo, A. (2020). Agile business model innovation in digital entrepreneurship: Lean startup approaches. *Journal of Business Research*, *110*, 519–537.

Goldammer, P., Annen, H., Stöckli, P. L., & Jonas, K. (2020). Careless responding in questionnaire measures: Detection, impact, and remedies. *The Leadership Quarterly*, *31*(4).

Grass, A., Backmann, J., & Hoegl, M. (2020). From empowerment dynamics to team adaptability: Exploring and conceptualizing the continuous agile team innovation process. *Journal of Product Innovation Management*, *37*(4), 324–351.

Guchait, P., & Hamilton, K. (2013). The temporal priority of team learning behaviors vs. shared mental models in service management teams. *International Journal of Hospitality Management*, *33*, 19–28.

Guinan, P. J., Cooprider, J. G., & Faraj, S. (1998). Enabling software development team performance during requirements definition: A behavioral versus technical approach. *Information Systems Research*, *9*(2), 101–125.

Gustavsson, T. (2016). Benefits of agile project management in a non-software development context: A literature review. In *Fifth International Scientific Conference on Project Management in the Baltic Countries, April 14-15* (pp. 114–124).

Hackman, J. R., & Oldham, G. R. (1980). *Work redesign*. Addison-Wesley.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis*. The Guilford Press.

Highsmith, J. (2004). *Agile project management: Creating innovative products*. Addison-Wesley.

Hoda, R., & Murugesan, L. K. (2016). Multi-level agile project management challenges: A self-organizing team perspective. *Journal of Systems and Software*, *117*, 245–257.

Hoda, R., Noble, J., & Marshall, S. (2012). Self-organizing roles on agile software development teams. *IEEE Transactions on Software Engineering*, *39*(3), 422–444.

Hoegl, M., & Parboteeah, K. P. (2006). Team reflexivity in innovative projects. *R&D Management*, *36*(2), 113–125.

Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology*, *27*(1), 99–114.

Hummel, M. (2014). State-of-the-art: A systematic literature review on agile information systems development. In *47th Hawaii International Conference on System Sciences* (pp. 4712–4721).

Ilgen, D. R. (1999). Teams embedded in organizations: Some implications. *American Psychologist*, *54*(2), 129–139.

Ilgen, D. R., Hollenbeck, J. R., Johnson, M., & Jundt, D. (2005). Teams in organizations: From input-process-output models to IMOI models. *Annual Review of Psychology*, *56*, 517–543.

Jiang, X., Flores, H. R., Leelawong, R., & Manz, C. C. (2016). The effect of team empowerment on team performance: A cross-cultural perspective on the mediating roles of knowledge sharing and intra-group conflict. *International Journal of Conflict Management*, *27*(1), 62–87.

Johnson, J. A. (2005). Ascertaining the validity of individual protocols from web-based personality inventories. *Journal of Research in Personality*, *39*(1), 103–129.

Kellermanns, F. W., Walter, J., Lechner, C., & Floyd, S. W. (2005). The lack of consensus about strategic consensus: Advancing theory and research. *Journal of Management*, *31*(5), 719–737.

Kirkman, B. L., & Rosen, B. (1999). Beyond self-management: Antecedents and consequences of team empowerment. *Academy of Management Journal*, *42*(1), 58–74.

Kirkman, B. L., Rosen, B., Tesluk, P. E., & Gibson, C. B. (2004). The impact of team empowerment on virtual team performance: The moderating role of face-to-face interaction. *Academy of Management Journal*, *47*(2), 175–192.

Klein, K. J., & Kozlowski, S. W. (2000a). From micro to meso: Critical steps in conceptualizing and conducting multilevel research. *Organizational Research Methods*, *3*(3), 211–236.

Klein, K. J., & Kozlowski, S. W. E. (2000b). *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions*. Jossey-Bass.

Klimoski, R., & Mohammed, S. (1994). Team mental model: Construct or metaphor? *Journal of Management*, *20*(2), 403–437.

Konradt, U., Otte, K.-P., Schippers, M. C., & Steenfatt, C. (2016). Reflexivity in teams: A review and new perspectives. *The Journal of Psychology*, *150*(2), 153–174.

Konradt, U., Schippers, M. C., Garbers, Y., & Steenfatt, C. (2015). Effects of guided reflexivity and team feedback on team performance improvement: The role of team regulatory processes and cognitive emergent states. *European Journal of Work and Organizational Psychology*, *24*(5), 777–795.

Laanti, M. (2013). Agile and Wellbeing — Stress, Empowerment, and Performance in Scrum and Kanban Teams. In *46th Hawaii International Conference on System Sciences* (pp. 4761–4770).

Laanti, M., Salo, O., & Abrahamsson, P. (2011). Agile methods rapidly replacing traditional methods at Nokia: A survey of opinions on agile transformation. *Information and Software Technology*, *53*(3), 276–290.

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods*, *9*(2), 202–220.

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, *11*(4), 815–852.

LePine, J. A., Piccolo, R. F., Jackson, C. L., Mathieu, J. E., & Saul, J. R. (2008). A meta-analysis of teamwork processes: Tests of a multidimensional model and relationships with team effectiveness criteria. *Personnel Psychology*, *61*(2), 273–307.

Lester, S. W., Meglino, B. M., & Korsgaard, M. A. (2002). The antecedents and consequences of group potency: A longitudinal investigation of newly formed work groups. *Academy of Management Journal*, *45*(2), 352–368.

Lewis, K. (2004). Knowledge and performance in knowledge-worker teams: A longitudinal study of transactive memory systems. *Management Science*, *50*(11), 1519–1533.

Lim, B.-C., & Klein, K. J. (2006). Team mental models and team performance: A field study of the effects of team mental model similarity and accuracy. *Journal of Organizational Behavior*, *27*(4), 403–418.

Locke, E. A. (1976). The Nature and Causes of Job Satisfaction. In *Handbook of Industrial and Organizational Psychology* (Vol. 1, pp. 1297–1349). Rand McNally.

Mafakheri, F., Nasiri, F., & Mousavi, M. (2008). Project agility assessment: An integrated decision analysis approach. *Production Planning & Control*, *19*(6), 567–576.

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *Academy of Management Review*, *26*(3), 356–376.

Marshall, C., & Rossmann, G. B. (2006). *Designing qualitative research* (4th ed.). Sage Publications.

Mathieu, J. E., Gilson, L. L., & Ruddy, T. M. (2006). Empowerment and team effectiveness: An empirical test of an integrated model. *Journal of Applied Psychology*, *91*(1), 97–108.

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Cannon-Bowers, J. A., & Salas, E. (2005). Scaling the quality of teammates' mental models: Equifinality and normative comparisons. *Journal of Organizational Behavior*, *26*(1), 37–56.

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology*, *85*(2), 273–283.

Mathieu, J. E., Hollenbeck, J. R., van Knippenberg, D., & Ilgen, D. R. (2017). A century of work teams in the Journal of Applied Psychology. *Journal of Applied Psychology*, *102*(3), 452–467.

Mathieu, J. E., Maynard, M. T., Rapp, T., & Gilson, L. (2008). Team effectiveness 1997-2007: A review of recent advancements and a glimpse into the future. *Journal of Management*, *34*(3), 410–476.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*(3), 709–734.

Maynard, M. T., & Gilson, L. L. (2014). The role of shared mental model development in understanding virtual team effectiveness. *Group & Organization Management*, *39*(1), 3–32.

Maynard, M. T., Mathieu, J. E., Gilson, L. L., O'Boyle Jr, E. H., & Cigularov, K. P. (2013). Drivers and outcomes of team psychological empowerment: A meta-analytic review and model test. *Organizational Psychology Review*, *3*(2), 101–137.

McHugh, O., Conboy, K., & Lang, M. (2011). Agile practices: The impact on trust in software project teams. *IEEE Software*, *29*(3), 71–76.

Meinecke, A. L., Hemshorn de Sanchez, C. S., Lehmann-Willenbrock, N., & Buengeler, C. (2019). Using state space grids for modeling temporal team dynamics. *Frontiers in Psychology*, *10*, 863.

Mertens, D. M. (2005). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods* (2nd ed.). Sage Publications.

Misra, S. C., Kumar, V., & Kumar, U. (2009). Identifying some important success factors in adopting agile software development practices. *Journal of Systems and Software*, *82*(11), 1869–1890.

Moe, N. B., Dingsøyr, T., & Dybå, T. (2008, March). Understanding Self-Organizing Teams in Agile Software Development. In *19th Australian Conference on Software Engineering* (pp. 76–85).

Moe, N. B., Dingsøyr, T., & Dybå, T. (2010). A teamwork model for understanding an agile team: A case study of a Scrum project. *Information and Software Technology*, *52*(5), 480–491.

Mohammed, S., Ferzandi, L., & Hamilton, K. (2010). Metaphor no more: A 15-year review of the team mental model construct. *Journal of Management*, *36*(4), 876–910.

Morgeson, F. P., Reider, M. H., & Campion, M. A. (2005). Selecting individuals in team settings: The importance of social skills, personality characteristics, and teamwork knowledge. *Personnel Psychology*, *58*(3), 583–611.

Mullen, B., & Copper, C. (1994). The relation between group cohesiveness and performance: An integration. *Psychological Bulletin*, *115*(2), 210–227.

Mullins, L. J. (2016). *Management & Organisational Behaviour* (11th ed.). Pearson.

Orasanu, J. (1994). Shared Problem Models and Crew Decision Making. In N. Johnston, N. McDonald, & R. Fuller (Eds.), *Aviation Psychology in Practice* (pp. 255–285). Ashgate Publishing, NASA.

Otte, K.-P., Knipfer, K., & Schippers, M. C. (2019). Team Reflection: A Catalyst of Team Development and the Attainment of Expertise. In P. Ward et al. (Eds.), *The Oxford Handbook of Expertise* (pp. 1000–1020). Oxford University Press.

Otte, K.-P., Konradt, U., Garbers, Y., & Schippers, M. C. (2017). Development and validation of the REMINT: A reflection measure for individuals and teams. *European Journal of Work and Organizational Psychology*, *26*(2), 299–313.

Papadakis, E., & Tsironis, L. (2018). Hybrid methods and practices associated with agile methods, method tailoring and delivery of projects in a non-software context. *Procedia Computer Science*, *138*, 739–746.

Qumer, A., & Henderson-Sellers, B. (2008). A framework to support the evaluation, adoption and improvement of agile methods in practice. *Journal of Systems and Software*, *81*(11), 1899–1919.

Rasker, P., Post, W., & Schraagen, J. (2000). Effects of two types of intra-team feedback on developing a shared mental model in Command & Control teams. *Ergonomics*, *43*(8), 1167–1189.

Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a "Big Five" in Teamwork? *Small Group Research*, *36*(5), 555–599.

Schippers, M. C., Homan, A. C., & Van Knippenberg, D. (2013). To reflect or not to reflect: Prior team performance as a boundary condition of the effects of reflexivity on learning and final team performance. *Journal of Organizational Behavior*, *34*(1), 6–23.

Schmidt, C., Kude, T., Heinzl, A., & Mithas, S. (2014). How Agile Practices Influence the Performance of Software Development Teams: The Role of Shared Mental Models and Backup. In *Proceedings of the 34th International Conference on Information Systems* (pp. 1–18).

Schwaber, K. (2004). *Agile Project Management With SCRUM*. Microsoft Press.

Seibert, S. E., Silver, S. R., & Randolph, W. A. (2004). Taking empowerment to the next level: A multiple-level model of empowerment, performance, and satisfaction. *Academy of Management Journal*, *47*(3), 332–349.

Seibert, S. E., Wang, G., & Courtright, S. H. (2011). Antecedents and consequences of psychological and team empowerment in organizations: A meta-analytic review. *Journal of Applied Psychology*, *96*(5), 981–1003.

Serrador, P., & Pinto, J. K. (2015). Does Agile work? — A quantitative analysis of agile project success. *International Journal of Project Management*, *33*(5), 1040–1051.

Sharp, H., & Robinson, H. (2010). Three 'C's of Agile Practice: Collaboration, Co-ordination and Communication. In T. Dingsøyr, T. Dybå, & N. B. Moe (Eds.), *Agile Software Development* (Vol. 34, pp. 61–85). Springer.

Sheffield, J., & Lemétayer, J. (2013). Factors associated with the software development agility of successful projects. *International Journal of Project Management*, *31*(3), 459–472.

Smith, K. G., Smith, K. A., Olian, J. D., Sims Jr, H. P., O'Bannon, D. P., & Scully, J. A. (1994). Top management team demography and process: The role of social integration and communication. *Administrative Science Quarterly*, *39*(3), 412–438.

Smith-Jentsch, K. A., Mathieu, J. E., & Kraiger, K. (2005). Investigating linear and interactive effects of shared mental models on safety and efficiency in a field setting. *Journal of Applied Psychology*, *90*(3), 523–535.

Spector, P. E. (1997). *Job Satisfaction: Application, Assessment, Causes, and Consequences*. Sage Publications.

Spreitzer, G. M. (1995). Psychological empowerment in the workplace: Dimensions, measurement, and validation. *Academy of Management Journal*, *38*(5), 1442–1465.

Spreitzer, G. M. (1996). Social structural characteristics of psychological empowerment. *Academy of Management Journal*, *39*(2), 483–504.

Spreitzer, G. M., Kizilos, M. A., & Nason, S. W. (1997). A dimensional analysis of the relationship between psychological empowerment and effectiveness satisfaction, and strain. *Journal of Management*, *23*(5), 679–704.

Spreitzer, G. M., Noble, D. S., Mishra, A. K., & Cooke, W. N. (1999). Predicting process improvement team performance in an automotive firm: Explicating the roles of trust and empowerment. In R. Wageman (Ed.), *Research on Managing Groups and Teams: Groups in Context* (Vol. 2, pp. 71–92).

Srivastava, A., Bartol, K. M., & Locke, E. A. (2006). Empowering leadership in management teams: Effects on knowledge sharing, efficacy, and performance. *Academy of Management Journal*, *49*(6), 1239–1251.

Stewart, G. L., & Barrick, M. R. (2000). Team structure and performance: Assessing the mediating role of intrateam process and the moderating role of task type. *Academy of Management Journal*, *43*(2), 135–148.

Stout, R. J., Cannon-Bowers, J. A., Salas, E., & Milanovich, D. M. (1999). Planning, shared mental models, and coordinated performance: An empirical link is established. *Human Factors*, *41*(1), 61–71.

Stray, V., Moe, N. B., & Hoda, R. (2018). Autonomous agile teams. In A. Aguiar (Ed.), *Proceedings of the 19th International Conference on Agile Software Development: Companion* (pp. 1–5).

Tessem, B. (2014). Individual empowerment of agile and non-agile software developers in small teams. *Information and Software Technology*, *56*(8), 873–889.

Tjosvold, D., Tang, M. M., & West, M. (2004). Reflexivity for team innovation in China: The contribution of goal interdependence. *Group & Organization Management*, *29*(5), 540–559.

Tuckman, B. W. (1965). Developmental sequence in small groups. *Psychological Bulletin*, *63*(6), 384–399.

Van Der Vegt, G., Emans, B., & Van De Vliert, E. (2000). Team members' affective responses to patterns of intragroup interdependence and job complexity. *Journal of Management*, *26*(4), 633–655.

Wageman, R., Gardner, H., & Mortensen, M. (2012). The changing ecology of teams: New directions for teams research. *Journal of Organizational Behavior*, *33*(3), 301–315.

West, M. A. (1996). Reflexivity and work group effectiveness: A conceptual integration. In *Handbook of Work Group Psychology* (pp. 555–579). Wiley.

West, M. A. (2012). *Effective teamwork: Practical lessons from organizational research*. Wiley.

Whitworth, E., & Biddle, R. (2007, August). The social nature of agile teams. In *Agile 2007 (agile 2007)* (pp. 26–36).

Williams, L., & Cockburn, A. (2003). Agile Software Development: It's about Feedback and Change. *IEEE Software*, *36*(6), 39–43.

Yang, M., Schloemer, H., Zhu, Z., Lin, Y., Chen, W., & Dong, N. (2020). Why and when team reflexivity contributes to team performance: A moderated mediation model. *Frontiers in Psychology*, *10*, 1–10.

Yu, X., & Petter, S. (2014). Understanding agile software development practices using shared mental models theory. *Information and Software Technology*, *56*(8), 911–921.

Junior Management Science 8(1) (2023) 148-162

Junior Management Science

journal homepage: www.jums.academy

Promoting Transparency or Harming Shareholders? U.S. Market Reaction to the Regulation of Proxy Advisors

Förderung der Transparenz oder Kosten für die Aktionäre? Beurteilung der Regulierung von Proxy Advisors anhand der U.S.-Marktreaktion

Julian Schneider

Ludwig-Maximilians-Universität München

Abstract

The SEC's decision to regulate proxy advisors triggers conflicting reactions. The intended increased transparency for shareholders is opposed to higher costs for obtaining information. This paper addresses the question of whether shareholders perceive the regulation of proxy advisors positively as an added value or negatively as a cost. An event study is used to examine market reactions to eleven events that influenced the likelihood of regulation adoption. The event study reveals a negative overall market reaction, suggesting that shareholders view regulation less as an added value through increased transparency and more as a detriment through added costs. Reactions to events that increase (decrease) the likelihood of adoption tend to be negative (positive). Moreover, market reactions are stronger the lower the share of institutional investors in a company. This paper contributes to the debate on proxy advisor regulation by being the first to provide empirical evidence on how shareholders respond to regulation.

Zusammenfassung

Der Beschluss der SEC zur Regulierung von Proxy Advisors löst gegensätzliche Reaktionen aus. Der beabsichtigten gesteigerten Transparenz für Aktionäre stehen höhere Kosten zur Informationsbeschaffung gegenüber. Die vorliegende Arbeit befasst sich mit der Frage, ob Aktionäre die Regulierung von Proxy Advisors positiv als Mehrwert oder negativ als Kosten auffassen. Mit Hilfe einer Ereignisstudie werden die Marktreaktionen auf elf Ereignisse untersucht, die die Wahrscheinlichkeit der Einführung der Regulierung beeinflusst haben. Die Ereignisstudie zeigt eine negative Gesamtmarktreaktion, die darauf schließen lässt, dass Aktionäre die Regulierung weniger als Mehrwert durch erhöhte Transparenz, sondern vielmehr als Nachteil durch zusätzliche Kosten bewerten. Die Reaktionen auf Ereignisse, die die Wahrscheinlichkeit der Einführung erhöhen (verringern), fallen tendenziell negativ (positiv) aus. Zudem sind die Marktreaktionen stärker, je geringer der Anteil institutioneller Investoren an einem Unternehmen ist. Diese Arbeit trägt zur Debatte über die Regulierung von Proxy Advisors bei, indem sie als Erste empirische Erkenntnisse darüber liefert, wie Aktionäre auf eine Regulierung reagieren.

Keywords: Proxy Advisors; Corporate Governance; Ereignisstudie; Informationsintermediäre.

1. Einleitung

Im Juli 2020 hat die Securities and Exchange Commission (SEC) in den USA beschlossen, Proxy Advisors (PAs) stärker zu regulieren (Securities and Exchange Commission (SEC), 2020b). PAs sind Unternehmen, die institutionelle Investoren und andere Marktteilnehmer zu ihrer Stimmvergabe bei Hauptversammlungen beraten und bei der administrativen Durchführung des Abstimmungsprozesses unterstützen. Als Intermediäre spielen sie eine wichtige Rolle für ihre Kunden, da institutionelle Investoren in der Regel eine Vielzahl an Stimmanteilen halten und große Informationsmengen zu verarbeiten haben (Securities and Exchange Commission (SEC), 2020a, S. 5-8). Ihre Funktion verleiht den PAs signifikanten Einfluss auf öffentliche Märkte (Shubber, 2020). Um potenzielle Interessenkonflikte sichtbarer zu machen und vollständigere Informationen für die Aktionäre

DOI: https://doi.org/10.5282/jums/v8i1pp148-162

vor der Stimmvergabe zu erhalten, sieht die neue Regulierung vor, dass PAs ihre Empfehlungen gleichzeitig ihren Kunden und auch den analysierten öffentlichen Unternehmen zur Verfügung stellen. Zusätzlich müssen sie ihren Kunden die Kommentare der Unternehmen bereitstellen. Kritiker bemängeln, dass die Regulierung dadurch zu höheren Kosten für Aktionäre führt und die Kontrolle der Macht der Unternehmen einschränkt (Kiernan, 2020). Es ist also fraglich, ob die Regulierung der PAs in den USA den von der SEC beabsichtigten Mehrwert für Aktionäre darstellt oder ihnen letztendlich, wie von Kritikern befürchtet, Entscheidungen erschwert und Kosten verursacht.

Ziel dieser Arbeit ist die Beantwortung der Frage, ob Aktionäre die Regulierung von PAs in den USA positiv als Mehrwert oder negativ als Kosten auffassen. Da die Regulierung bis Dezember 2021 nicht verpflichtend ist, kann zum jetzigen Zeitpunkt noch nicht beurteilt werden, welchen Effekt sie tatsächlich bewirken wird. Anhand der Marktreaktionen in den USA können jedoch die Erwartungen der Aktionäre auf die Regulierung mit Hilfe einer Ereignisstudie beobachtet werden. Zunächst werden das Geschäftsmodell, der Einfluss auf den U.S.-Markt sowie die Kritik an PAs näher beleuchtet und der neue Regulierungsbeschluss der SEC erläutert. Anschließend werden signifikante Ereignisse im Zusammenhang mit der Verabschiedung der Regulierung identifiziert. Empfinden Aktionäre die Regulierung als Nachteil, wird erwartet, dass Ereignisse, die die Wahrscheinlichkeit der Verabschiedung erhöhen (verringern), zu einer negativen (positiven) Marktreaktion führen. Die Reaktionen des U.S.-Marktes werden mit Market Adjusted Returns (MAR) unmittelbar um die identifizierten Ereignisse analysiert, um herauszufinden, ob sie signifikant unterschiedlich von Null sind.

2. Proxy Advisors am U.S.-Markt

2.1. Geschäftsmodell von Proxy Advisors

Das Geschäftsmodell von PAs besteht primär darin, im Vorfeld von Hauptversammlungen börsennotierter Unternehmen Analysen für Aktionäre zu erstellen und Abstimmungsempfehlungen zu geben, um ihnen die Entscheidungsfindung zu erleichtern (European Securities and Markets Authority (ESMA), 2012, S. 9). Ihre Empfehlungen betreffen Schlüsselthemen wie die Beurteilung von Aktionärsanträgen, die Wahl und Vergütung des Vorstands, die Qualität der Corporate Governance, Nachhaltigkeit oder angestrebte Firmenübernahmen. Ihre Berichte verkaufen sie auf Basis von Abonnements vor allem an institutionelle Investoren wie Investmentfonds, Pensionsfonds, Versicherungen oder Vermögensverwalter (Government Accountability Office (GAO), 2016, S. 2; Hayne & Vance, 2019, S. 973).

In den USA erhalten Aktionäre der meisten börsennotierten Unternehmen ein Abstimmungsrecht, wenn sie Stammaktien kaufen. Nur wenige von ihnen geben ihre Stimme persönlich bei Hauptversammlungen ab, sondern wählen per „Proxy Vote“, indem sie einen Stimmrechtsvertreter bevollmächtigen, online oder per Brief abstimmen (Spatt, 2021, S. 136). Vor Hauptversammlungen versenden die Unternehmen daher Informationsmaterialien, „Proxy Statements“, über abzustimmende Punkte an ihre Aktionäre, damit diese eine fundierte Entscheidung treffen können. Mit über 70% Anteil sind institutionelle Investoren die größten Aktionäre börsennotierter Unternehmen in den USA (ProxyPulse, 2020, S. 2). Ihre Kunden haben sie zur Ausübung ihres Proxy Voting Rechts ermächtigt. Durch Beteiligungen an zahlreichen börsennotierten Unternehmen sind institutionelle Investoren für hunderte bis tausende Hauptversammlungen stimmberechtigt, die in der Regel in einem engen Zeitfenster von wenigen Monaten abgehalten werden (Securities and Exchange Commission (SEC), 2020a, S. 7-8). Sie stehen daher in kurzer Zeit sowohl vor der Herausforderung, unabhängige Abstimmungsentscheidungen zum Wohl ihrer Kunden zu treffen, als auch den Abstimmungsvorgang administrativ zu organisieren. Aus diesem Grund greifen sie auf die Dienstleistungen von PAs zurück. PAs haben standardisierte Recherche- und Analyseprozesse entwickelt, um öffentlich verfügbare Informationen zu aggregieren und zu verarbeiten. Sie analysieren unter anderem die Proxy Statements, die die Unternehmen vor Hauptversammlungen an ihre Aktionäre versenden, um darauf basierend Abstimmungsempfehlungen zu entwickeln. Ihre Einnahmen generieren sie zum einen dadurch, dass sie Investoren individuelle, auf ihre Richtlinien zugeschnittene Berichte verkaufen. Zum anderen verkaufen PAs standardisierte Berichte auf Basis ihrer eigenen „Benchmark Policy“ an teils hunderte Kunden. Eine Benchmark Policy[1] ist eine Richtlinie, die aufführt, anhand welcher Aspekte PAs Unternehmen vergleichen und ihre Abstimmungsempfehlungen entwickeln. Die Richtlinien werden regelmäßig mit Hilfe von eigenen Erhebungen und dem Input verschiedener Marktteilnehmer, insbesondere großer institutioneller Investoren, angepasst (Hayne & Vance, 2019, S. 1005-1006). Die Berichte werden von institutionellen Investoren in verschiedenen Weisen genutzt, unter anderem als Alternative oder Ergänzung für ihre eigenen Analysen (Securities and Exchange Commission (SEC), 2020a, S. 7). Durch die Informationen erhalten sie ein tieferes Verständnis der Abstimmungspunkte, um eine fundierte Entscheidung treffen zu können und die eigenen Ressourcen effizienter zu nutzen. Der Zweck von PAs ist es, institutionellen Investoren die rechtzeitige und informierte Ausübung ihrer Stimmrechte zu erleichtern (European Securities and Markets Authority (ESMA), 2012, S. 9). Durch das Sammeln, Verarbeiten und Verbreiten von Informationen, fungieren sie als wichtige Intermediäre, die die Informations- und Kontrollkosten ihrer Kunden reduzieren (Hitz & Lehmann, 2018). Zusätzlich zur Informationsbereitstellung bieten PAs auch Unterstützung bei der administrativen Durchführung der Stimmabgabe an. In der Regel stellen sie eine elektronische Plattform zur Verfügung, die den Abstimmungsprozess effizienter gestaltet (Securities and Exchange Commission (SEC), 2020a, S. 8). In den vergangenen Jahren haben sich einige PAs ein weiteres Geschäftsfeld erschlossen. Neben der Beratung von Aktionären, bieten sie

[1] Synonym wird oft auch „Voting Policy“ verwendet.

auch Beratungsleistungen für die börsennotierten Unternehmen selbst zu Themen wie Vorstandsvergütung, Corporate Governance und Nachhaltigkeit an. Mit ihrer Hilfe versuchen die Unternehmen die Abstimmungsergebnisse bei Hauptversammlungen zu verbessern (Hayne & Vance, 2019, S. 976; Securities and Exchange Commission (SEC), 2020a, S. 51).

2.2. Ursprung und Einfluss auf den U.S.-Markt

Der Ursprung von PAs geht zurück auf die 1980er Jahre, als Institutional Shareholder Services, Inc. (ISS), als eines der ersten Unternehmen dieser Art, in den USA gegründet wurde. In den 2000er Jahren kamen in kurzer Zeit weitere hinzu, unter anderem Glass, Lewis & Co., LLC (GL) (European Securities and Markets Authority (ESMA), 2012, S. 10). Der schnelle Anstieg von PAs wurde durch eine neue Richtlinie begünstigt, die die SEC 2003 verabschiedete. Richtlinie 206(4)-6 des Investment Advisers Act von 1940 (Advisers Act) besagt, dass Anlageberater[2] eine Treuepflicht („Fiduciary Duty") gegenüber ihren Kunden haben. Gemäß ihrer Treuepflicht müssen Anlageberater Richtlinien zur Stimmrechtsvertretung entwickeln, die frei von Interessenkonflikten und angemessen gestaltet sind, um sicherzustellen, dass sie die Proxy Votes im besten Interesse ihrer Kunden abstimmen. Die neuen Anforderungen der Richtlinie 206(4)-6 stellten eine erhebliche wirtschaftliche Belastung für Anlageberater dar, die die nötigen Ressourcen dafür nicht aufwenden wollten (Placenti, 2019). Gleichzeitig ebneten sie den Weg für PAs. Basierend auf einer Leitlinie der SEC-Mitarbeiter („Staff Guidance"), war es Anlageberatern möglich, ihre Treuepflicht, insbesondere die Vermeidung von Interessenkonflikten, zu erfüllen, indem sie ihre Proxy Votes auf die Empfehlungen von PAs stützten. Das hatte zur Folge, dass immer mehr institutionelle Investoren ihre Recherchen und Abstimmungsentscheidungen an PAs auslagerten und somit deren Wachstum stark beschleunigten (Government Accountability Office (GAO), 2016, S. 14-15). Diese Praxis manifestierte sich im folgenden Jahr durch die Veröffentlichung der sogenannten No-Action Letters. In zwei öffentlichen Briefen stellte die SEC 2004 klar, dass Anlageberater die Analyse von Proxy Statements und den Proxy Voting Prozess an unabhängige Dritte wie PAs auslagern können. Außerdem müssten die Anlageberater nicht in jedem einzelnen Fall prüfen, ob zwischen den PAs und den börsennotierten Unternehmen Interessenkonflikte vorliegen, sondern könnten sich auf die allgemeinen Konfliktverfahren der PAs verlassen. Das ermöglichte es diesen wiederum, sowohl institutionellen Investoren als auch börsennotierten Unternehmen Beratungsdienstleistungen anzubieten (Rose & Walker, 2020, S. 7). Obwohl es die Absicht der SEC war, ein größeres Engagement institutioneller Investoren im Proxy Voting Prozess anzustoßen, hatte die Richtlinie somit den gegenteiligen Effekt (Doyle, 2018, S. 6). Heute wird der U.S.-Markt von zwei PAs dominiert, die zusammen einen Marktanteil von 97% haben (Hayne & Vance, 2019, S. 973). ISS, das 2020 mehrheitlich in den Besitz der Deutsche Börse (2020) übergegangen ist, ist mit 60% Marktanteil der größte PA in den USA, gefolgt von GL mit 37% (Spatt, 2021, S. 137). Mit ihrem Wachstum stieg über die Jahre jedoch auch die Kritik an PAs. Nicht nur auf Grund der hohen Konzentration der Marktanteile, sondern vor allem wegen der Abhängigkeit institutioneller Investoren wuchs die Angst vor zu viel Einfluss durch blindes Vertrauen in die Empfehlungen von PAs. Bereits 2007 führte das U.S. Government Accountability Office (GAO) in einem Bericht an, dass kleinere institutionelle Investoren nicht die Kapazitäten hätten, um aufwändige Recherchen zu betreiben und daher dazu geneigt seien, sich komplett auf die Empfehlungen von PAs zu verlassen (Government Accountability Office (GAO), 2016, S. 2). Diese Praxis wird als „Robo-Voting" bezeichnet (Doyle, 2018, S. 6). Gemäß einer Studie des American Council for Capital Formation (ACCF) folgen 175 institutionelle Investoren, die zusammen ein Vermögen von mehr als 5 Billionen US-Dollar repräsentieren, bei Abstimmungen zu 95% den Empfehlungen von PAs (Doyle, 2018, S. 5). Gleichzeitig liegt die Abstimmungsquote bei institutionellen Investoren bei 92% im Gegensatz zu Kleinanlegern mit 28%, was ihr Gewicht bei Proxy Votes verstärkt (ProxyPulse, 2020, S. 2).

Der größte Teil akademischer Literatur über PAs befasst sich mit der Beantwortung der Frage, welchen Einfluss sie auf den U.S.-Markt haben. Hayne und Vance (2019, S. 975, 1003) untersuchten beispielsweise, ob PAs indirekt als Intermediäre über ihre Empfehlungen Einfluss auf die Vorstandsvergütung[3] von börsennotierten Unternehmen haben oder gar direkten Einfluss als de facto Standardsetzer ausüben. Hintergrund dieser Untersuchung ist die Kritik, dass die hohe Übereinstimmung der Stimmvergabe institutioneller Investoren mit den Empfehlungen von PAs die Unternehmen dazu drängt, ihre Vergütungsstruktur an den öffentlich einsehbaren Benchmark Policies der PAs auszurichten, um negative Ergebnisse zu vermeiden. Dadurch würden die Richtlinien von PAs zu „Standards" für Unternehmen und sie selbst zu Standardsetzern. Hayne und Vance kommen zu dem Schluss, dass PAs sowohl indirekt als auch direkt Einfluss auf die Vorstandsvergütung börsennotierter Unternehmen ausüben. Zu dem gleichen Schluss kommen Larcker, McCall und Ormazabal (2015, S. 173). Sie kritisieren, dass Unternehmen durch PAs dazu verleitet werden, Entscheidungen zu treffen, die den Shareholder Value verringern. Malenko und Shen (2016) zeigen, dass eine negative Empfehlung von ISS in Bezug auf die Vorstandsvergütung zu einem um 25 Prozentpunkte niedrigeren Abstimmungsergebnis führen. Ertimur, Ferri und Oesch (2013) berichten sogar, dass mehr als die Hälfte aller Unternehmen als Reaktion auf eine negative Empfehlung im Anschluss an die Abstimmung ihre

[2] Als Anlageberater, auf Englisch „Investment Advisers" oder auch „Asset Manager", sind unter anderem institutionelle Investoren wie Investmentfonds und ETFs zu verstehen, siehe Spatt (2021).

[3] Seit dem Dodd-Frank Act von 2010 können Aktionäre über die Vergütung des Vorstands abstimmen („say on pay"). Das Ergebnis ist zwar nicht bindend, aber es sendet ein wichtiges Signal über die Zufriedenheit der Aktionäre mit der Vergütung und Leistung des Vorstands aus, siehe Larcker, Tayan und Copland (2018).

Vergütungsstruktur geändert haben. Alexander, Chen, Seppi und Spatt (2010) und Brav, Jiang, Li und Pinnington (2019) kommen ebenfalls zu der Erkenntnis, dass PAs einen signifikanten Einfluss auf den Ausgang von Proxy Votes in den USA haben. Diese Beobachtung beschränkt sich nicht nur auf den U.S.-Markt. Hitz und Lehmann (2018, S. 740) haben die Auswirkungen der Empfehlungen von ISS und GL auf Unternehmen in Europa untersucht. Auch sie folgern, dass PAs eine wichtige ökonomische Rolle eingenommen haben, die zwar weniger ausgeprägt ist als in den USA, aber dennoch dazu führt, dass ihre Empfehlungen signifikante Marktreaktionen auslösen.

2.3. Kritik an Proxy Advisors

Wie in Abschnitt 2.2. erläutert, ist der Einfluss von PAs auf Proxy Votes und die Corporate Governance von börsennotierten Unternehmen in den USA empirisch belegt. Angesichts dessen, der hohen Marktkonzentration auf praktisch nur zwei Firmen und der Abhängigkeit vieler institutioneller Investoren, stehen PAs stark in der Kritik. Vor allem börsennotierte Unternehmen fürchten negative Auswirkungen. 2019 forderten 319 von ihnen in einem von der Nasdaq Börse organisierten Brief die SEC dazu auf, PAs stärker zu kontrollieren (Griggs, 2019). Die Kritik umfasst in der Regel zwei wesentliche Punkte: Potenzielle Interessenkonflikte und die Qualität der Empfehlungen.

Interessenkonflikte entstehen vor allem dann, wenn PAs sowohl Corporate Governance Beratungsleistungen für Unternehmen anbieten als auch Abstimmungsempfehlungen über diese Unternehmen an institutionelle Investoren verkaufen. Die SEC gab bei der Veröffentlichung eines Konzeptpapiers an, dass Interessenkonflikte die vielleicht meistgenannte Sorge in Bezug auf PAs sind (Securities and Exchange Commission (SEC), 2010, S. 116). Diese Konstellation bezieht sich allerdings ausschließlich auf ISS und ihr Tochterunternehmen ISS Corporate Solutions, da GL keine Corporate Governance Beratung anbietet. Einige Unternehmen deuteten an, dass sie sich unter Druck gesetzt fühlten ISS' Beratung in Anspruch zu nehmen, um negative Empfehlungen bei ihren Proxy Votes zu verhindern. Im Umkehrschluss wird befürchtet, dass Unternehmen, die Corporate Governance Beratung erhalten, von ISS bevorzugt werden könnten (Government Accountability Office (GAO), 2007, S. 10). Ein weiteres Feld für Interessenkonflikte ist die Beteiligungsstruktur einzelner PAs. Bis zu ihrer Übernahme durch die Deutsche Börse AG befand sich ISS zu 100% in Besitz des institutionellen Investors Genstar Capital, LLC. Mittlerweile beläuft sich der Anteil von Genstar nur noch auf 20% (Deutsche Börse, 2020). GL wurde ebenfalls kürzlich verkauft, ist jedoch komplett im Besitz eines institutionellen Investors geblieben. Ihre Anteile wurden an den Private Equity Investor Peloton Capital Management verkauft (McElhaney, 2021). Der Interessenkonflikt wird hier in Bezug auf Aktionärsanträge („Shareholder Proposals") gesehen, die auf den Hauptversammlungen zur Wahl stehen. Es besteht die Möglichkeit, dass die Investoren, die PAs besitzen, Aktionärsanträge einreichen und von ihren Tochterunternehmen Empfehlungen zu ihren Gunsten erhalten. Außerdem ist es möglich, dass Besitzer von PAs oder Mitglieder ihrer Führungsebene eine erhebliche Beteiligung an Unternehmen haben, zu denen Empfehlungen gegeben werden (Government Accountability Office (GAO), 2007, S. 11). Generell geht ein potenzieller Interessenkonflikt aus dem Geschäftsmodell von PAs hervor. Dadurch, dass sie von ihnen bezahlt werden, besteht von Natur aus ein Anreiz mehr Empfehlungen zu Gunsten von Aktionärsanträgen institutioneller Investoren zu geben, um die Geschäftsbeziehungen mit ihnen aufrecht zu erhalten (Cappucci, 2019, S. 21). In einer Rede äußerte der SEC Commissioner Elad Roisman (2020a) zudem Bedenken über eine Praxis, die „Social Arbitrage" genannt wird. Institutionelle Investoren würden die Abstimmungsempfehlungen von PAs beeinflussen, sodass bei Proxy Votes beispielsweise gegen ein unbeliebtes Vorstandsmitglied gewählt wird. Dies sei eine Möglichkeit, indirekt gegen das Unternehmen zu stimmen, ohne gleichzeitig die Geschäftsbeziehungen zu gefährden. Obwohl einige PAs bereits Maßnahmen zur Offenlegung von Interessenkonflikten umsetzen, fordern Kritiker einheitlichere Regeln. Ohne sie könne es zu inkonsistenten oder unzureichenden Offenlegungen kommen (Business Roundtable, 2020). Die Empfehlungen von ISS enthalten zum Beispiel eine Legende, die andeutet, dass das betreffende Unternehmen Kunde ihrer Tochtergesellschaft ISS Corporate Solutions sein könnte (Securities and Exchange Commission (SEC), 2019a, S. 92). Dieser Hinweis bietet jedoch keine Gewissheit, ob dem Unternehmen tatsächlich Beratungsdienstleistungen angeboten werden bzw. in welchem Umfang die Geschäftsbeziehung besteht.

Der zweite Hauptkritikpunkt an PAs thematisiert die Qualität ihrer Empfehlungen. Angesichts der hohen Übereinstimmung zwischen ihren Empfehlungen und den Abstimmungsentscheidungen institutioneller Investoren (Doyle, 2018, S. 5) erhalten Nachlässigkeiten oder Fehler bei der Recherche umso mehr Gewicht. Oft wird kritisiert, dass sie falsche Peergroups auswählen oder fehlerhafte Kalkulationen durchführen (Cappucci, 2019, S. 19). Außerdem wird bemängelt, dass PAs ihre Recherchen in Billiglohnländer wie die Philippinen auslagern und ihre Qualitätskontrollen inadäquat seien. In einer Umfrage des COEC (2011, S. 10) gaben 53% der Befragten an, dass PAs einen oder mehrere Fehler in ihren Berichten über die Vorstandsvergütung von Unternehmen gemacht haben. Der Unternehmensanwalt Placenti (2018, S. 11) stellte in einer Untersuchung von 107 Abstimmungsempfehlungen zu 97 verschiedenen Unternehmen aus den Jahren 2016 bis 2018 faktische, analytische oder andere ernsthafte Fehler fest. Das SEC Investor Advisory Commitee (IAC) (2020, S. 5) entgegnete jedoch, dass der Anteil faktischer Fehler angesichts tausender Abstimmungen bei unter einem Prozent läge. Da PAs innerhalb weniger Monate tausende Unternehmen analysieren müssen – allein ISS hat über 4.600 Kunden (Institutional Shareholder Services (ISS), 2021) – verwenden sie aus Zeit- und Kostengründen ihre standardisierten Benchmark Policies. Ihnen wird deshalb ein „one-size-fits-all"-Ansatz vorgeworfen, der sich nur auf Kosteneffizienz fokussiere. Ein solcher Ansatz sei aber nicht in der Lage,

auf die Besonderheiten der Unternehmen einzugehen und ihre Corporate Governance angemessen zu beurteilen (Rose & Walker, 2020, S. 10). Wenn institutionelle Investoren PAs auf Basis von Kosten anstatt der Qualität auswählten, bestehe die Gefahr, dass sie keinen Anreiz mehr haben, tiefgreifende firmenspezifische Untersuchungen durchzuführen (Larcker et al., 2015, S. 175). Zudem wird die fehlende Transparenz bei der Entwicklung ihrer Benchmark Policies kritisiert. Diese seien oft auf fehlerhafte akademische Studien gestützt und es lasse sich nicht prüfen, ob sie richtig auf die jeweiligen Unternehmen angewendet werden (Rose & Walker, 2020, S. 10). Kritiker fordern aus diesen Gründen, dass Unternehmen, zu denen PAs ihre Empfehlungen aussprechen, in die Entwicklung der Benchmark Policies eingebunden werden sollen, um zu verhindern, dass PAs zu de facto Standardsetzern werden (Griggs, 2019). Außerdem sollen die Unternehmen eine Möglichkeit bekommen, die Empfehlungen rechtzeitig vor den Proxy Votes einsehen und kommentieren zu können, um potenzielle Fehler aufzudecken, bevor sie die Abstimmung beeinflussen. Zwar geben ISS und GL Unternehmen im Fall einer negativen Empfehlung bereits die Möglichkeit den Entwurf vorab einzusehen und Korrekturen anzumerken, jedoch erhalten nicht alle Unternehmen diese Möglichkeit oder sie erhalten sie nicht rechtzeitig. Eine Umfrage des ACCF von 100 Unternehmen ergab, dass 37% (84%) von ihnen keine Möglichkeit zur Vorabeinsicht von ISS (GL) erhalten haben. 85% erhielten weniger als 72 Stunden Antwortzeit von ISS (Placenti, 2018, S. 7-8).

SEC Commissioner Roisman (2020a) ist der Meinung, dass die Leidtragenden dieser Kritikpunkte die „Main Street"-Investoren sind. Kleinanleger, die mehr und mehr große Teile ihrer Ersparnisse in Fonds investieren, um von den Vorzügen institutioneller Investoren zu profitieren, aber gleichzeitig auch die Kosten ihrer Abstimmungsentscheidungen zu tragen hätten. Er sieht daher eine stärkere Regulierung von PAs und ihrer Empfehlungen als notwendig an, um zu gewährleisten, dass institutionelle Investoren im besten Sinne ihrer Kunden abstimmen.

2.4. Regulierung von Proxy Advisors

Bis zum Jahr 2018 wurden PAs nur durch den Advisers Act von 1940 reguliert. Durch verschiedene Ausnahmen konnten sie frei bestimmen, ob sie sich selbst als Anlageberater bei der SEC registrieren wollen. Damit einher gingen zwar zusätzliche regulatorische Anforderungen, jedoch wurden sie in Bezug auf PAs nur wenig durchgesetzt (COEC, 2011, S. 11). Die Registrierung als Anlageberater hatte dadurch eher symbolischen als regulatorischen Charakter für PAs. Von den beiden Marktführern ist nur ISS als Anlageberater geführt, GL hat darauf verzichtet (Government Accountability Office (GAO), 2016, S. 10). Eine wichtige Erleichterung für sie waren die No-Action Letters von 2004, die es anderen Anlageberatern ermöglichten ihre Treuepflicht zu erfüllen, indem sie ihre Abstimmungsentscheidungen auf die Empfehlungen von PAs stützten (siehe Abschnitt 2.2.). Der dadurch gewonnene Einfluss und die lauter werdende Kritik sorgte dafür, dass die Debatte um eine stärkere Regulierung von PAs immer hitziger wurde.

Die SEC reagierte auf die Kritik mit einem über 10 Jahre andauernden Regulierungsprozess. Bereits 2010 veröffentlichte sie ein Konzeptpapier zum U.S. Proxy Voting System, in dem sie unter anderem Kommentare zur Rolle von PAs einforderte. Die SEC thematisierte potenzielle Interessenkonflikte und mangelnde Genauigkeit und Transparenz bei der Entwicklung von Abstimmungsempfehlungen (Securities and Exchange Commission (SEC), 2010, S. 114-126). 2013 veranstalteten sie einen Roundtable[4], bei dem der Einsatz von PAs durch institutionelle Investoren diskutiert wurde (Securities and Exchange Commission (SEC), 2013). Im Juni 2014 veröffentlichten die Mitarbeiter der SEC ein Rundschreiben („Staff Legal Bulletin"), das Anlageberatern Leitlinien zu ihren Treuepflichten bei der Beauftragung von PAs gab (Securities and Exchange Commission (SEC), 2014). Darin wurde ihnen geraten, Richtlinien und Prozesse zu entwickeln, um die Arbeit der PAs und potenzielle Interessenkonflikte zu kontrollieren. Dennoch blieben die No-Action Letters von 2004 weiter in Kraft, sodass das Rundschreiben nur einen geringen Effekt auf den Einsatz von PAs hatte (Posner, 2018). Es wurde jedoch als Hinweis interpretiert, dass Anlageberater sich nicht kritiklos auf die Abstimmungsempfehlungen von PAs verlassen sollen (Placenti, 2019). Da die Forderungen nach einer stärkeren Regulierung unvermindert blieben, hielt die SEC im November 2018 erneut einen Roundtable ab. Im Vorfeld dieser Veranstaltung nahmen die Mitarbeiter im September 2018 offiziell die No-Action Letters zurück, um die Diskussion über eine Überarbeitung der Regulierung zu erleichtern (Securities and Exchange Commission (SEC), 2018b). Die Rücknahme der Letters stellte in Frage, inwieweit sich Anlageberater auf die Empfehlungen von PAs verlassen können und signalisierte einen grundlegenden Wandel in der Haltung der SEC gegenüber PAs (Placenti, 2019). Im August 2019 veröffentlichte sie zwei Leitlinien („Guidance"), die die Kritik auf zwei Wegen aufgriffen. Eine Guidance richtete sich an Anlageberater als Kunden von PAs, die andere richtete sich direkt an PAs.

In der ersten Guidance stellte die SEC die Verantwortlichkeiten von Anlageberatern bei Proxy Votes im Zusammenhang mit der Erfüllung ihrer Treuepflicht klar. Dabei legte sie spezielles Augenmerk auf Anlageberater, die PAs beschäftigen (Securities and Exchange Commission (SEC), 2019b, S. 8). Die Guidance erhielt durch die Rücknahme der No-Action Letters eine stärkere Wirkung als zuvor im September 2018, da Anlageberater sich nicht mehr in gleichem Maß auf die Empfehlungen von PAs verlassen konnten, um ihrer Treuepflicht nachzukommen. Die SEC regte sie somit zu einer insgesamt kritischeren Grundhaltung an. Sie sollten unter anderem mögliche Interessenkonflikte bereits bei der Beauftragung identifizieren, die Abstimmungsempfehlungen vermehrt auf Fehler oder Unvollständigkeiten untersuchen und die Kapazität und Kompetenz der PAs stärker kontrollieren.

[4] Ein Roundtable ist ein Treffen der SEC mit Interessengruppen zur Diskussion eines bestimmtes Themas.

Die SEC stellte auch klar, dass Anlageberater nicht jede Möglichkeit in Anspruch nehmen müssen, für ihre Kunden abzustimmen. Das bezieht sich auf Fälle, in denen dies explizit mit ihren Kunden vereinbart wurde und auf Fälle, in denen es im besten Interesse des Kunden ist, nicht an der Abstimmung teilzunehmen, zum Beispiel wenn die Kosten den Nutzen übersteigen (Securities and Exchange Commission (SEC), 2019b, S. 17-24). Dadurch reduzierte sie zugleich die Notwendigkeit, sich auf die Beratung von PAs zu verlassen (Spatt, 2021, S. 153). Mit dieser Guidance legte die SEC eine hohe Messlatte für Anlageberater, ihre Treuepflichten zu erfüllen und ging damit auf die Kritik am großen Einfluss von PAs durch Robo-Voting ein (Placenti, 2019).

In der zweiten Guidance stellte die SEC klar, dass PAs nach ihrer Ansicht den Proxy Rules unterliegen. Proxy Rules sind Richtlinien, die im Securities Exchange Act von 1934 (Exchange Act) festgelegt worden sind und den Abstimmungsprozess regulieren. Bis zu diesem Zeitpunkt fanden die Proxy Rules auf PAs keine Anwendung, da die herrschende Meinung war, dass ihre Abstimmungsempfehlungen keine „Solicitation“ im Sinne der Richtlinie 14a-1(l) der Proxy Rules darstellen. Mit der neuen Guidance änderte sich diese Ansicht. Der Begriff „Solicitation“ wird in Richtlinie 14a-1(l) nicht genau definiert, die SEC versteht ihn in Verbindung mit der Auslegung vorangegangener Jahre jedoch als eine Kommunikation mit Wertpapierinhabern, die nach vernünftiger Beurteilung darauf abzielt, die Beschaffung, Zurückhaltung oder den Widerruf seiner Stimmabgabe zu bewirken (Securities and Exchange Commission (SEC), 2019c, S. 4). Das Government Accountability Office (GAO) (2016, S. 6) definiert einen Proxy Solicitor als eine Firma, die börsennotierten Unternehmen dabei hilft mit Aktionären zu kommunizieren, um ihre Stimmen zu sichern. Man kann eine Solicitation also als Kommunikationsprozess mit Aktionären bezeichnen, der darauf abzielt, die Stimmvergabe des Aktionärs zu beeinflussen. Die SEC interpretiert Abstimmungsempfehlungen von PAs als Solicitation, weil sie darauf abzielten, ihre Kunden bei der Abstimmungsentscheidung zu unterstützen, wodurch eine Beeinflussung entstünde. Dies gelte auch, wenn die Kunden nicht ihren Empfehlungen folgten (Securities and Exchange Commission (SEC), 2019c, S. 9-10). Entscheidend ist, dass eine Solicitation strenger Offenlegungspflichten durch die Proxy Rules unterliegt. Demnach muss die Kommunikation mit dem Aktionär über ein aufwändiges Proxy Statement erfolgen, es sei denn, man greift auf die Ausnahmen aus Richtlinie 14a-2(b) zurück (Government Accountability Office (GAO), 2016, S. 8-9). Richtlinie 14a-2(b)(1) befreit Personen, die nicht versuchen als Stimmrechtsvertreter für einen Wertpapierinhaber zu handeln oder anderweitig eine Abstimmung einzureichen oder anzufordern. 14a-2(b)(3) befreit Personen, die andere Personen zu ihrer Stimmvergabe beraten, wenn sie bestimmte Kriterien erfüllen. Diese Kriterien umfassen vor allem die Offenlegung signifikanter Beziehungen mit dem Unternehmen, zu dem die Abstimmungsempfehlung gegeben wird und dass der Berater nur von dem Kunden vergütet wird, der die Beratung erhält. Die SEC behielt sich in ihrer Guidance außerdem vor, die Kriterien von Paragraf 14a-2(b)(3) zukünftig auszuweiten (Securities and Exchange Commission (SEC), 2019c, S. 3). Die Feststellung, dass die Proxy Rules auf PAs zutreffen, bedeutet zusätzlich, dass sie den Anti-Fraud Bestimmungen von Richtlinie 14a-9 unterliegen, die sie für ihre Empfehlungen haftbar machen können. In 14a-9 steht, dass Solicitations keine Aussagen enthalten dürfen, die falsch oder irreführend in Bezug auf eine wesentliche Tatsache sind. In ihrer Guidance merkte die SEC an, dass sich die Richtlinie auch auf Meinungen, Begründungen, Empfehlungen oder Überzeugungen erstreckt und somit PAs explizit miteinschließt. Sie empfiehlt in diesen Fällen eine Offenlegung von zugrundeliegenden Fakten, Annahmen, Einschränkungen oder anderer Informationen, damit sie keine Bedenken im Sinne der Anti-Fraud Bestimmungen erwecken. PAs sollten daher in Erwägung ziehen, die Methodik zu erläutern, die sie zur Formulierung ihrer Abstimmungsempfehlung verwendet haben, zusätzlich andere Quellen als die öffentlich verfügbaren Unternehmensinformationen zu veröffentlichen und wesentliche Interessenkonflikte offenzulegen. Sollten zur Entwicklung der Empfehlungen Peergroups verwendet werden, sollten auch die dafür ausgewählten Unternehmen benannt werden (Securities and Exchange Commission (SEC), 2019c, S. 12-13). Mit der Auslegung als Solicitation griff die SEC die Kritik an der mangelnden Qualität der Empfehlungen von PAs und ihren unzureichend offengelegten Interessenkonflikten auf. Durch die Anwendung der Proxy Rules sollten sie zu sorgfältigerer Arbeit angehalten und ein transparenterer Einblick in die Entwicklung ihrer Empfehlungen gewährt werden.

Im November 2019 erhöhte die SEC den Druck auf PAs noch einmal. Sie veröffentlichte einen Vorschlag zur Änderung der Proxy Rules, ebenfalls um die Genauigkeit und Transparenz von Abstimmungsempfehlungen zu verbessern (Securities and Exchange Commission (SEC), 2019a). Der Vorschlag stellte eine Verankerung der zweiten Guidance vom August 2019 in den bestehenden Richtlinien des Exchange Acts dar und enthielt weitreichende Anpassungen. Bezüglich Richtlinie 14a-1(l) schlug die SEC eine „Kodifizierung“ der Interpretation von Solicitation vor, indem explizit die Stimmrechtsberatung mit aufgenommen werden sollte (Securities and Exchange Commission (SEC), 2019a, S. 17). Um weiterhin Gebrauch von den Ausnahmen in Richtlinie 14a-2(b) machen zu können, wurde vorgeschlagen die Kriterien, die dafür erfüllt werden müssen, zu erweitern und in den speziell für PAs neu geschaffenen Paragrafen 14a-2(b)(9) aufzunehmen. Nach Ansicht der SEC waren die bisherigen Vorschriften zur Offenlegung von Interessenkonflikten durch PAs zu ungenau und sollten daher erheblich detaillierter in Paragraf 14a-2(b)(9)(i) festgehalten werden. Zusätzlich wurden zwei neue Kriterien vorgeschlagen. Gemäß Paragraf 14a-2(b)(9)(ii) sollten PAs ihre Abstimmungsempfehlungen den betroffenen Unternehmen rechtzeitig für eine Vorabeinsicht zur Verfügung stellen, bevor sie an ihre Kunden verschickt werden. Dadurch hätten Unternehmen die Möglichkeit faktische, methodische oder anderweitige Mängel aufzudecken und den PAs mitzuteilen, um die Qualität ihrer Empfehlungen zu erhöhen. Das zweite neue Kri-

terium in Paragraf 14a-2(b)(9)(iii) sah darüber hinaus vor, dass die Abstimmungsempfehlungen auf Nachfrage einen Hyperlink enthalten müssen, der zu einer Stellungnahme der Unternehmen bezüglich der Empfehlungen führt. Durch diese Maßnahme sollte den Kunden ein effizienterer Zugriff auf die Ansichten der Unternehmen und möglichst vollständige Informationen gewährleistet werden (Securities and Exchange Commission (SEC), 2019a, S. 30-54). Zuletzt enthielt der Vorschlag der SEC auch eine Anpassung der Anti-Fraud Bestimmungen in Richtlinie 14a-9. Dort ist eine Liste mit Beispielen irreführender Sachverhalte aufgeführt. Diese Liste sollte um Informationen ergänzt werden, die PAs veröffentlichen sollen, um eine Haftung für ihre Empfehlungen zu vermeiden (Securities and Exchange Commission (SEC), 2019a, S. 68-72). Die vorgeschlagenen Änderungen der Proxy Rules stellten einen einschneidenden Regulierungsversuch dar. Vor allem die Anpassung der Ausnahmen in Richtlinie 14a-2(b) wurde kontrovers diskutiert. Durch die Möglichkeit zur Vorabeinsichtnahme und die Bereitstellung von Antworten wäre es Unternehmen möglich, gegen eine negative Abstimmungsempfehlung der PAs vorzugehen, argumentierten Kritiker (Placenti, 2019). Im Anschluss an die Veröffentlichung der Vorschläge wurde allen Parteien Zeit gegeben, Kommentare bei der SEC einzureichen, um Bedenken oder Anregungen zu äußern.

Im Juli 2020 verkündete die SEC schließlich die finale Richtlinie. Sie stellt im Wesentlichen eine Umsetzung der im November 2019 vorgeschlagenen Anpassungen dar, mit wenigen, aber wichtigen Änderungen. Der umstrittene Vorschlag, dass PAs ihre Empfehlungen vor dem Versenden an ihre Kunden zuerst zur Kontrolle den betreffenden Unternehmen zukommen lassen müssen, wurde entscheidend geändert. Laut SEC wurden viele kritische Kommentare seit der Veröffentlichung des Vorschlags berücksichtigt, sodass der finale Paragraf 14a-2(b)(9)(ii)(A) letztendlich abgeschwächt wurde. Wenn PAs Empfehlungen an ihre Kunden geben, müssen sie diese nicht zwingend vorab den Unternehmen zur Kommentierung zur Verfügung stellen, sondern können dies auch gleichzeitig tun (Securities and Exchange Commission (SEC), 2020a, S. 90-91). PAs sind jedoch gemäß dem neuen Paragrafen 14a-2(b)(9)(ii)(B) dazu verpflichtet, ihren Kunden Zugriff auf die Antworten der Unternehmen zu ihren Empfehlungen zu gewähren (S. 91). Der Vorschlag dazu vom November 2019 wurde unverändert übernommen. Allerdings hat die SEC dahingehend eine Ausnahme geschaffen. PAs werden von beiden Anforderungen in Paragraf 14a-2(b)(9)(ii) befreit, wenn ihre Abstimmungsempfehlungen auf den spezifischen Benchmark Policies ihrer Kunden beruhen. Diese Ausnahme wurde in Paragraf 14a-2(b)(9)(v) erfasst (S. 113). Neben der finalen Überarbeitung der Proxy Rules hat die SEC im Juli 2020 auch eine Erweiterung ihrer Guidance für Anlageberater vom August 2019 veröffentlicht. Darin macht sie Anlageberater darauf aufmerksam, dass sie, wenn sie PAs beschäftigen, die Antworten der Unternehmen ebenfalls berücksichtigen müssen, um ihre Treuepflicht zu erfüllen (Securities and Exchange Commission (SEC), 2020c, S. 4-5). Der finale Regulierungsbeschluss adressiert die Kritik an PAs auf verschiedene Weise. Durch die Anwendung der Proxy Rules unterliegen PAs verstärkter Offenlegungspflichten, die Interessenkonflikte aufdecken sollen. Sie müssen mehr mit den Unternehmen, die Gegenstand ihrer Empfehlungen sind, interagieren und vollständigere und transparentere Informationen an ihre Kunden geben. Durch die Anti-Fraud Bestimmungen wird ihre Haftbarkeit für falsche oder irreführende Informationen erhöht, um eine höhere Qualität ihrer Arbeit zu bezwecken. Gleichzeitig werden Anlageberater durch die zusätzliche Guidance auf ihre Treuepflicht aufmerksam gemacht, um unkritisches Vertrauen in die Empfehlungen von PAs zu verhindern und ihren Einfluss auf Abstimmungen zu beschränken. Obwohl die finale Richtlinie weniger drastisch als der Vorschlag vom November 2019 ausgefallen ist, sieht die SEC in ihr ein wirksames Mittel, um „Main Street"-Investoren besser zu schützen (Roisman, 2020b). Die beschlossene Regulierung ist jedoch nicht unumstritten. ISS verklagte die SEC bereits im Oktober 2019 wegen der vorhergehenden Guidance (Retelny, 2019). Nach einer zwischenzeitlichen Unterbrechung (Posner, 2020) wurde das Verfahren nach der Verkündung der finalen Regulierung erneut aufgenommen (Flook, 2020). Das Ergebnis ist zum aktuellen Zeitpunkt noch ungewiss.

Neben der SEC hat auch der U.S.-Kongress seit 2016 mehrfach die Initiative ergriffen, um PAs zu regulieren. Mit dem Corporate Governance Reform and Transparency Act (US-Congress, 2016, 2017, 2019) wurden drei Gesetzentwürfe im Repräsentantenhaus vorgestellt, die unter anderem vorsahen, dass sich PAs bei der SEC registrieren müssen. Einzig der Entwurf von 2017 wurde genehmigt, fand jedoch anschließend im Senat keine Berücksichtigung. Ein weiterer Versuch war der Corporate Governance Fairness Act US-Congress (2018), der zusätzlich zu einer Registrierung von PAs auch eine regelmäßige Untersuchung durch die SEC anordnete. Auch dieser wurde nicht genehmigt. Zuletzt hat das U.S. Department of Labor (DOL) im Dezember 2020 eine neue Richtlinie für Pensions- und Gesundheitspläne der Privatwirtschaft verabschiedet, die ähnlich wie die SEC die Treuepflichten von Anlageberatern hervorhebt und ihnen strengere Überwachungspflichten bei der Beschäftigung von PAs auferlegt (Department of Labor (DOL), 2020, S. 81668). Die Richtlinie wurde allerdings im März 2021 nachträglich bis zu ihrer Überarbeitung gestoppt (Department of Labor (DOL), 2021). In Europa beschäftigt sich die European Securities and Markets Authority (ESMA) seit 2011 mit der Regulierung von PAs. Mit der Shareholder Rights Directive wurde 2017 in dieser Hinsicht eine Vorgabe für EU-Mitgliedsstaaten beschlossen, die besagt, dass PAs einem Verhaltenskodex unterliegen, über die Anwendung dieses Kodex berichten und bestimmte Schlüsselinformationen in Bezug auf die Erstellung ihrer Empfehlungen, sowie alle tatsächlichen oder potenziellen Interessenkonflikte offenlegen müssen (2017/828/EC, 2017).

Wie in Abschnitt 2.2. beschrieben, gibt es bereits viele Arbeiten, die Erkenntnisse über den Einfluss und die Rolle von PAs liefern. Sie stellen die empirische Basis dar, auf

der eine Regulierung beruht.[5] Allerdings ist Literatur, die eine Regulierung von PAs in den Vordergrund stellt, bisher sehr limitiert. In der Regel stellen diese Arbeiten den Status Quo der Regulierung zum Zeitpunkt ihres Erscheinens dar und geben Hinweise für potenzielle Anpassungen. Belinfanti (2009) konzentriert sich diesbezüglich auf die USA, während Klöhn und Schwarz (2013) einen marktübergreifenden Ansatz wählen. Gallego Córcoles (2016) richtet ihren Blick auf Europa. Balp (2017) bezieht sich ebenfalls auf Europa, allerdings stellt ihre Arbeit eine kritische Auseinandersetzung mit der Shareholder Rights Directive der ESMA dar. Eckstein (2015) beschreibt negative Effekte der Regulierung von PAs, die im Zusammenhang mit einer Erwartungslücke entstehen können. Daher schlägt er Lösungen vor, wie die Erwartungslücke bei der Ausarbeitung neuer Richtlinien verkleinert werden kann. Rose und Walker (2020) beziehen sich auf die jüngst verabschiedete Regulierung durch die SEC in den USA, indem sie den Regulierungsprozess seit 2010, Kommentare und die verabschiedeten Richtlinien zusammenfassen. Der Konsens dieser Arbeiten ist, dass ihre Autoren eine Regulierung von PAs für nötig erachten. Wie die Reaktion des Marktes auf eine Regulierung ausfällt, ist bisher nicht bekannt. Das mag vor allem daran liegen, dass PAs die längste Zeit von einer direkten Regulierung verschont geblieben sind.

3. Empirische Analyse

3.1. Potenzielle Effekte der Regulierung

Naturgemäß spalten sich die Ansichten über die Effekte einer Regulierung in zwei Lager. Das Lager der Profiteure, in diesem Fall börsennotierte Unternehmen und vermeintlich Kleinanleger, sowie das Lager der von der Regulierung Betroffenen, in diesem Fall PAs und institutionelle Investoren. Die von der SEC intendierten und von den Unternehmen größtenteils unterstützten (Business Roundtable, 2020) positiven Effekte auf Proxy Votes und Kleinanleger werden in Abschnitt 2.4. beschrieben. PAs und institutionelle Investoren hingegen prognostizieren vor allem negative Effekte für Aktionäre. Commissioner Herren Lee (2020), die einzige Gegenstimme innerhalb der SEC, fasst die Bedenken, die PAs und institutionelle Investoren in Kommentaren an die SEC gerichtet haben, am Tag der Verkündung der neuen Richtlinie zusammen. Sie bezeichnet die Regulierung als ungerechtfertigt, unerwünscht und nicht umsetzbar. Sie sei ungerechtfertigt, weil ihre Beweggründe nicht beweisbar seien. Laut IAC liegt die Fehlerrate in den Empfehlungen von PAs bei unter einem Prozent (Investor Advisory Commitee (IAC), 2020, S. 5). Zudem sei die Regulierung unerwünscht, wie eine Auswertung der eingereichten Kommentare zeigt. Demnach unterstützen 46 Kommentare die Bestimmung zur Bereitstellung der Empfehlungen an Unternehmen und zum Einbezug ihrer Antworten, während 159 Kommentare dagegen sind. Grund dafür sei, dass diese Neuerungen den Investoren Verzögerungen und Mehrkosten bereiten, obwohl die Regulierung eigentlich darauf abziele Investoren zu schützen. Die neuen Bestimmungen seien nicht umsetzbar, weil der Einbezug der Antworten dazu führe, dass institutionelle Investoren dazu verpflichtet wären zum einen mehr Informationen zu verarbeiten und zum anderen nicht klar sei, wie lange sie vor der Abstimmung auf die Antworten warten müssten (Herren Lee, 2020). Das Council of Institutional Investors (CII) befürchtet außerdem, dass PAs im Zuge der erhöhten Haftbarkeit ihre Unabhängigkeit verlieren und ihre Empfehlungen mehr an der Sicht der Unternehmen orientieren könnten. Zusätzlich würde die Richtlinie höhere Kosten für PAs bei der Erarbeitung ihrer Empfehlungen verursachen, die an die institutionellen Investoren und somit letztendlich an die „Main Street"-Investoren weitergegeben werden würden. Ein PA schätzt die tatsächlichen Kosten auf das 240-fache der Schätzung der SEC (Council of Institutional Investors (CII), 2020, S. 6; 2020b). Das IAC sieht dadurch vor allem kleine und mittelgroße Anlageberater gefährdet, die stärker auf PAs angewiesen sind, weil sie nicht über die Mittel für eine komplett eigenständige Analyse verfügen (Investor Advisory Commitee (IAC), 2020, S. 2). Diese Aussage deckt sich mit den Erkenntnissen, zu denen das GAO bereits 2007 gekommen ist (Government Accountability Office (GAO), 2016, S. 2), deutet aber auch an, dass es unter institutionellen Investoren unterschiedliche Positionen gibt. BlackRock, der weltweit größte Vermögensverwalter, hat bereits 2018 in einem Kommentar zum Roundtable der SEC seine Unterstützung zur Regulierung von PAs ausgesprochen (Novick & Cameron, 2018). BlackRock selbst zählt jedoch zu der Gruppe großer institutioneller Investoren, die über die nötigen Ressourcen verfügen, um ihre Abstimmungsentscheidungen selbst zu erarbeiten.

SEC Commissioner Herren Lee (2020) fasst zusammen, dass die finale Regulierung die Stimmabgabe für Aktionäre schwerer und kostenintensiver gestalte, wodurch sie sich weniger Gehör verschaffen und das Management der Unternehmen weniger zur Verantwortung ziehen könnten. Das CII sieht in der Regulierung das Ergebnis jahrelanger Lobbyarbeit börsennotierter Unternehmen (Council of Institutional Investors (CII), 2020). Ihre Intention sei weniger die Maximierung des Shareholder-Value, sondern vielmehr geringerer Widerstand bei Proxy Votes. Denn für institutionelle Investoren sei die Alternative zur Verwendung von PA-Empfehlungen nicht vermehrt eigene Analysen durchzuführen, sondern sich mehr auf die Vorschläge des Managements zu verlassen (Cappucci, 2019, S. 33).

3.2. Hypothesenherleitung

Die Regulierung von PAs soll dazu führen, dass institutionelle Investoren transparentere, genauere und vollständigere Empfehlungen erhalten, um sicherzustellen, dass sie im besten Interesse ihrer Kunden, den „Main Street"-Investoren, abstimmen. Gleichzeitig soll dadurch das Vertrauen in die Märkte gestärkt werden (Roisman, 2020b). Basierend auf

[5] Hitz und Lehmann (2018, S. 713) adressieren ihre Arbeit zum Beispiel explizit an die europäischen Regulatoren.

den in Abschnitt 3.1. genannten Effekten für institutionelle Investoren ist jedoch fraglich, ob sich der von der SEC beabsichtigte Mehrwert für Kleinanleger auch in der U.S.-Marktreaktion widerspiegelt. Institutionelle Investoren scheinen von der Regulierung durch erhöhten Zeitdruck und höhere Kosten größtenteils negativ beeinflusst zu werden. Zugleich sind sie jedoch mit einem Anteil von über 70% mit Abstand die größte Aktionärsgruppe börsennotierter U.S.-Unternehmen (ProxyPulse, 2020, S. 2). Aus diesem Grund werden folgende Hypothesen aufgestellt. Da institutionelle Investoren den größten Markteinfluss haben und die Kosten ihren Mehrwert überwiegen, wird erwartet, dass die Regulierung von PAs zu einer insgesamt negativen Gesamtreaktion des U.S.-Marktes führt. Somit wird erwartet, dass Ereignisse, die die Wahrscheinlichkeit der Regulierung erhöhen (verringern), zu einer negativen (positiven) Marktreaktion führen.

3.3. Identifikation von relevanten Ereignissen

In Tabelle 1 werden elf Ereignisse zwischen Dezember 2017 und September 2020 identifiziert, von denen erwartet wird, dass sie die Wahrscheinlichkeit der Einführung der Regulierung von PAs beeinflussen. Um sie zu identifizieren, wurde die Webseite der SEC hinsichtlich Pressemitteilungen, Reden, öffentlichen Stellungnahmen und Veranstaltungen, sowie die Webseite des U.S.-Kongress nach Gesetzentwürfen und Anhörungen in Bezug auf die Regulierung von PAs untersucht. Zusätzlich wurden die Nachrichten bekannter Zeitschriften und Nachrichtenagenturen wie die Financial Times, New York Times, Wall Street Journal und Reuters ausgewertet. Auf dieser Basis konnten ursprünglich 30 Ereignisse ermittelt werden (Anhang A), die auf elf reduziert wurden. Ereignisse, die keine neuen Informationen enthielten, wurden für die finale Auswahl nicht berücksichtigt. Ebenso beschränkt sich die Auswahl, bis auf eine Ausnahme, auf Ereignisse im Zusammenhang mit der SEC. Grund dafür ist, dass nur bei der SEC ernst zu nehmende Regulierungsbestrebungen zu erkennen waren.

Wie bereits in Abschnitt 2.4. erklärt wurde, hat die SEC seit 2010 vereinzelt Maßnahmen eingeleitet, die sich mit einer möglichen Regulierung von PAs auseinandersetzten. Diese waren bis 2018 jedoch zeitlich zerstreut und deuteten weniger auf eine tatsächliche Einschränkung von PAs hin. Ein Beispiel dafür ist das Staff Rundschreiben im Juni 2014, das zwar die Treuepflichten von Anlageberatern hervorhob, aber keinerlei rechtliche Bindung hatte und durch die noch immer geltenden No-Action Letters auch keine Konsequenzen befürchten ließ (Securities and Exchange Commission (SEC), 2014). Das änderte sich allerdings in dem Zeitraum ab 2018 bis Ende 2020, in dem auch eine deutliche Verdichtung der Ereignisse zu erkennen ist. Daher wurden die Handlungen der SEC vor 2018 ebenfalls nicht berücksichtigt. Tabelle 1 enthält außerdem für jedes Ereignis eine Einschätzung, ob es die Wahrscheinlichkeit der Regulierung erhöht oder verringert.

Das erste Ereignis im Dezember 2017 ist die Verabschiedung des US-Congress (2017) durch das House of Representatives. Obwohl im U.S.-Kongress mehrere Gesetzentwürfe zur Regulierung von PAs vorgestellt wurden, ist dies der Einzige, der die Zustimmung einer der beiden Kammern gewinnen konnte. Er wurde zwar später nicht vom Senat bestätigt, dennoch stellt er ein wichtiges Ereignis dar, denn darin wird die SEC gleichzeitig aufgefordert, die No-Action Letters von 2004 zurückzunehmen (S. 19). Es wird daher angenommen, dass dieses Ereignis die Wahrscheinlichkeit einer Regulierung erhöht. Das zweite Ereignis am 30. Juli 2018 ist die Ankündigung des SEC Staff Roundtables zum Proxy Voting Prozess. Darin richtete der SEC Vorsitzende Jay Clayton einige Vorschläge für Agendapunkte an seine Mitarbeiter. Unter anderem regte er an, das regulatorische Umfeld von PAs und die mögliche Aufhebung früherer Staff Guidance zu thematisieren (Clayton, 2018). Da Clayton damit indirekt eine Regulierung andeutete, wird auch dieses Ereignis als die Wahrscheinlichkeit erhöhend eingeordnet. In seiner Aussage kann bereits ein Hinweis auf das dritte Ereignis gedeutet werden. Am 13. September 2018 verkündeten die Mitarbeiter der SEC nämlich, dass die No-Action Letters von 2004 zurückgenommen werden (Securities and Exchange Commission (SEC), 2018b). Wie in Abschnitt 2.4. beschrieben, stellte diese Maßnahme einen tiefen Eingriff in die Grundlage des großen Einflusses von PAs dar, denn die No-Action Letters garantierten Anlageberatern die Erfüllung ihrer Treuepflichten durch die Beschäftigung von PAs. Es wird deswegen angenommen, dass das dritte Ereignis die Wahrscheinlichkeit der Regulierung erhöht. Außerdem kann angenommen werden, dass dadurch die Aufmerksamkeit auf den bevorstehenden Roundtable noch einmal erhöht wurde. Am 8. November 2018 verkündete die SEC schließlich die Teilnehmer und Agenda. Es zeigte sich, dass die Mitarbeiter den Vorschlag des Vorsitzenden Clayton aufgegriffen und PAs auf die Agenda gesetzt haben. Zudem wurden als Teilnehmer die Geschäftsführer von ISS und GL verkündet (Securities and Exchange Commission (SEC), 2018a). Dieses Ereignis stellt eine Bestätigung des Regulierungsvorhabens dar, das wiederum als die Wahrscheinlichkeit der Einführung erhöhend eingeordnet wird. Ereignis Nummer fünf ist der eigentliche Roundtable am 15. November 2018. In ihren Eröffnungsreden erläutern die SEC Commissioner Roisman (2018) und Stein (2018) wesentliche Bedenken in Bezug auf PAs und ziehen ein regulatorisches Eingreifen in Betracht, um den Problemen zu begegnen. Es wird angenommen, dass das Ereignis die Wahrscheinlichkeit der Regulierung erhöht. Das sechste Ereignis ist die Veröffentlichung der Guidance der SEC am 21. August 2019. Die Anwendung der Bestimmungen der Proxy Rules auf PAs und die Betonung der Treuepflicht von Anlageberatern stellen bedeutende Einschränkungen für PAs dar und legen das Fundament für die finale Regulierung. Das Ereignis wird daher erneut als Erhöhung der Wahrscheinlichkeit eingeordnet. Die Guidance löste großen Widerstand bei ISS aus, die in ihr eine Unterdrückung ihrer geschützten Meinungsäußerung und eine Gefahr für die Aktualität und Unabhängigkeit ihrer Beratung sahen. Ereignis Nummer sieben stellt die Einreichung der Klage von ISS gegen die August Guidance der SEC am 31. Oktober 2019 dar. Ihre Klage basiert auf drei Punkten. Erstens argumentiert ISS, dass ihre Abstimmungs-

Tabelle 1: Ereignisse und vorhergesagte Marktreaktion auf die Wahrscheinlichkeit der Regulierung von PAs

Ereignis	Beschreibung	Wahrscheinlichkeit der Einführung	Vorhergesagte Reaktion, wenn Nutzen >Kosten (Nutzen <Kosten)
(1) 20. Dezember 2017	House of Representatives beschließt Corporate Governance Reform and Transparency Act of 2017 (H.R. 4015)	Erhöhend	+ (-)
(2) 30. Juli 2018	SEC kündigt Roundtable zum Proxy Process an	Erhöhend	+ (-)
(3) 13. September 2018	SEC Staff zieht No-Action Letters von 2004 zurück	Erhöhend	+ (-)
(4) 08. November 2018	SEC kündigt Agenda und Teilnehmer des Roundtables an	Erhöhend	+ (-)
(5) 15. November 2018	SEC Roundtable zum Proxy Process	Erhöhend	+ (-)
(6) 21. August 2019	SEC Guidance für Investment Advisers und zur Anwendung der Proxy Rules auf PAs	Erhöhend	+ (-)
(7) 31. Oktober 2019	ISS klagt gegen die SEC Guidance vom 21. August 2019	Verringernd	- (+)
(8) 05. November 2019	SEC schlägt Änderung der Proxy Rules vor	Erhöhend	+ (-)
(9) 30. Januar 2020	SEC beantragt Unterbrechung des Gerichtsverfahrens mit ISS und setzt August Guidance aus	Verringernd	- (+)
(10) 22. Juli 2020	SEC beschließt Änderung der Proxy Rules	Erhöhend	+ (-)
(11) 18. September 2020	ISS setzt Klage gegen SEC fort	Verringernd	- (+)

empfehlungen keine Solicitation darstellten, weshalb es außerhalb der Befugnis der SEC läge, sie danach zu regulieren. Zweitens stelle die Guidance eine substanzielle Richtlinie dar und deshalb sei nach dem Administrative Procedure Act eine Mitteilungs- und Kommentierungsphase vorgeschrieben. Ihr dritter Punkt ist, dass die Guidance willkürlich und unberechenbar sei, weil sie eine wesentliche Änderung ihrer Beratung bedeute, die SEC aber behaupte, sie hätte ihre Position nicht verändert (Institutional Shareholder Services (ISS), 2019). Das gerichtliche Vorgehen von ISS wird als Verringerung der Wahrscheinlichkeit der Regulierung angesehen. Am 5. November 2019 veröffentlichte die SEC ihren Vorschlag zur Änderung der Proxy Rules, der PAs unter anderem dazu verpflichten sollte, ihre Abstimmungsempfehlungen zur Vorabeinsicht an die betreffenden börsennotierten Unternehmen zu senden (Securities and Exchange Commission (SEC), 2019a). Er setzte die PAs weiter unter Druck und wird als Erhöhung der Wahrscheinlichkeit der Regulierung eingeordnet. Als neuntes Ereignis wurde der Antrag der SEC auf Unterbrechung des Gerichtsverfahrens mit ISS am 30. Januar 2020 identifiziert. Im Zuge der bevorstehenden finalen Regulierung deutete die SEC an, dass sie wesentliche Punkte der Klage von ISS bei der Erarbeitung der neuen Richtlinie berücksichtigen würde und bis zu ihrer Verkündung auf die Durchsetzung der Guidance vom August 2019 verzichte (Posner, 2020). Weil die SEC damit ein Entgegenkommen auf die Forderungen der PAs andeutet, wird das Ereignis als die Wahrscheinlichkeit verringernd eingeordnet. Ereignis Nummer zehn ist der Beschluss zur Änderung der Proxy Rules am 22. Juli 2020. Wie in Abschnitt 2.4. näher erläutert, ist die finale Richtlinie weniger drastisch als die vorgeschlagenen Änderungen, was vor allem daran liegt, dass PAs ihre Empfehlungen nicht mehr vorab an die börsennotierten Unternehmen schicken müssen. Dennoch setzt sie große Teile des Vorschlags in einer prinzipienbasierten Weise um (Securities and Exchange Commission (SEC), 2020b). Das Ereignis erhöht die Wahrscheinlichkeit der Regulierung. Als elftes Ereignis wurde die Fortsetzung der Klage von ISS gegen die SEC am 18. September 2020 identifiziert. Die ursprüngliche Klage wurde um die neue Richtlinie erweitert. Als Gründe dafür gibt ISS an, dass die SEC nicht klarmachen konnte, warum die Regulierung notwendig sei und warum der regulatorische Rahmen zuvor unter dem Advisers Act nicht ausreichend gewesen wäre. Zusätzlich würde die neue Richtlinie ihr Recht auf Meinungsfreiheit verletzen, da sie gezwungen seien, die Meinungen der Unternehmen mit einzubeziehen (Flook, 2020). Es wird angenommen, dass die erneute Klage von ISS die Wahrscheinlichkeit der Regulierung verringert, da nicht ausgeschlossen werden kann, dass das Gericht zu ihren Gunsten entscheidet.

3.4. Datenbasis und Methodik

Da sich die Regulierung von PAs potenziell auf die Marktreaktionen aller börsennotierter U.S.-Unternehmen auswirken kann, wird als Ausgangspunkt ein CRSP Datensatz mit Beobachtungen von allen verfügbaren börsennotierten U.S.-Unternehmen an den Ereignistagen im Zeitraum von 2017 bis 2020 verwendet. Tabelle 2 fasst die Bearbeitung des Datensatzes zusammen.

Zunächst enthält der Datensatz 83.070 Beobachtungen von 9.021 Unternehmen zu jeweils mindestens einem der elf identifizierten Ereignisse. Es werden 32 Beobachtungen von Unternehmen ausgeschlossen, die keine Returns aufweisen. Um die Gesamtmarktreaktion zu testen, werden 3-Tages-Fenster (Event-Windows) um die Ereignistage gebildet. Die Verkündung von Quartalsergebnissen innerhalb des Event-

Tabelle 2: Bearbeitung des Datensatzes

Selektionskriterium	Beobachtungen	Firmen
Alle U.S.-Unternehmen mit Daten an Ereignistagen 2017-2020	83.070	9.021
abzüglich Beobachtungen ohne Return	(32)	(2)
abzüglich Beobachtungen mit Verkündung der Quartalsergebnisse im Event-Window	(4.212)	(2)
Finaler Datensatz für t-statistic	78.826	9.017

Windows stellt ein firmenspezifisches Störereignis (Confounding Event) dar. Es ist wahrscheinlich, dass die Marktreaktion auf die Quartalsergebnisse eine Reaktion auf die Regulierung von PAs überdeckt, daher werden die betroffenen Event-Windows, insgesamt 4.212 Beobachtungen, entfernt. Die drei Returns pro Event-Window werden aufsummiert, um anschließend für den kompletten U.S.-Markt einen Cumulative Return (CR_{USA}) über drei Tage pro Ereignis zu erhalten. Da die Regulierung potenziell alle börsennotierten U.S.-Unternehmen beeinflusst, wird als Vergleich ein globaler Index verwendet. Der DJSTOXX GLOBAL 1800 von Datastream enthält jeweils die 600 größten amerikanischen, europäischen und asiatischen (Asia-Pacific) Unternehmen. Die 600 amerikanischen Unternehmen werden von dem Index abgezogen (DJSTOXX GLOBAL 1800 ex America). Auch für den DJSTOXX 1800 werden die CR über Event-Windows von drei Tagen pro Ereignis berechnet und von den CR_{USA} abgezogen. Aus dieser Marktbereinigung resultiert der Cumulative Market Adjusted Return ($CMAR_{USA}$) für den U.S.-Markt für jedes Ereignis. Um testen zu können, ob die Gesamtmarktreaktion auf die Regulierung signifikant ausfällt, wird der $CMAR_{USA}$ bei Ereignissen, die die Wahrscheinlichkeit der Regulierung verringern (Ereignisse 7, 9 und 11), mit -1 multipliziert ($ACMAR_{USA}$). Durch diesen Schritt wird die Interpretation der aggregierten Ereignisse erleichtert, denn es wird verhindert, dass sich die Effekte der einzelnen Ereignisse mit verringernder oder erhöhender Wahrscheinlichkeit gegenseitig aufheben (Joos & Leung, 2013). Es entsteht ein Datensatz mit 78.826 Beobachtungen von 9.017 Unternehmen. Um herauszufinden, wie der U.S.-Markt auf die Regulierung von PAs reagiert, werden zwei t-Tests durchgeführt. Der erste untersucht, ob sich der Mittelwert der elf $ACMAR_{USA}$ signifikant von Null unterscheidet (t-statistic vs. 0). Der zweite t-Test (t-statistic vs. Non-Events) untersucht, ob sich der mittlere Return an Ereignistagen signifikant von dem mittleren Return an Tagen ohne Ereignis (Non-Events) im Zeitraum von 2017 bis 2020 unterscheidet. Die CMAR der Non-Events wurden ähnlich wie zuvor über ein 3-Tages-Fenster gebildet und durch den Vergleichsindex bereinigt.

3.5. Ergebnisse

Tabelle 3 fasst die Ergebnisse der Gesamtmarktreaktionen auf alle elf Ereignisse zusammen. Es lässt sich feststellen, dass der mittlere Return aller Ereignisse negativ ist (-0,1%) und sich signifikant von Null auf dem 1%-Niveau (t-statistic vs. 0 = -5,5211) unterscheidet. Dies deutet darauf hin, dass U.S.-Aktionäre die Regulierung von PAs insgesamt als Nachteil oder Kosten wahrnehmen und bestätigt die in Abschnitt 3.2. formulierte Hypothese. Der mittlere Return unterscheidet sich zudem signifikant von dem mittleren Return der Non-Events (t-statistic vs. Non-Events = 8,7987).

Es wurde erwartet, dass Ereignisse, die die Wahrscheinlichkeit der Einführung der Regulierung erhöhen, auf Grund der hohen Anteile institutioneller Investoren zu einer negativen Reaktion des Marktes führen. Betrachtet man die $CMAR_{USA}$ pro Ereignis, lässt sich diese Annahme teilweise bestätigen. Die Ankündigung des Roundtables zum Proxy Process am 30. Juli 2018 (-1%) und die Zurücknahme der No-Action Letters am 13. September 2018 (-1,2%) riefen die größten Reaktionen der Aktionäre hervor. Sie erhöhten die Wahrscheinlichkeit einer Regulierung und resultierten wie erwartet in negativen Returns. Allerdings bestätigen sich die Erwartungen nicht bei allen Ereignissen. Am 15. November 2018, dem Tag, an dem der Roundtable abgehalten wurde, ist eine signifikant positive Marktreaktion zu erkennen, obwohl angenommen wurde, dass das Ereignis die Wahrscheinlichkeit der Regulierung erhöht. Dieser Fall ist zugleich ein Beispiel für die Limitationen dieser Studie. Es kann nicht garantiert werden, dass die Einschätzung des Einflusses der Ereignisse auf die Wahrscheinlichkeit der Regulierung zutreffend ist. So wurde der Roundtable zwar im Bestreben einer Regulierung abgehalten, allerdings sahen die Aktionäre in den Diskussionen bei der Veranstaltung möglicherweise einen Hinweis darauf, dass die Regulierung nicht oder weniger strikt durchgesetzt wird. Am 22. Juli 2020 hingegen ist keine signifikante Reaktion festzustellen. Ein möglicher Grund dafür wäre, dass an diesem Tag zwar die Richtlinie zur Regulierung der PAs final beschlossen wurde, aber gleichzeitig weniger drastisch ausfiel, als ursprünglich befürchtet. Überraschend ist dagegen, dass es auf die Veröffentlichung der Guidance am 21. August 2019 und die anschließende Klage von ISS am 31. Oktober 2019 ebenfalls keine signifikanten Reaktionen gab.

Die Ereignisse Nummer eins und neun riefen signifikant gegensätzliche Marktreaktionen als erwartet hervor. Es wird angenommen, dass diese durch exogene Confounding Events verursacht wurden. Um auszuschließen, dass die beobachteten Reaktionen durch andere Ereignisse als die im Zusammenhang mit der Regulierung identifizierten verursacht wurden, wurden verschiedene Nachrichtenquellen überprüft. Mit Hilfe von Google News, der Financial Times, dem Wall Street Journal und der New York Times konnten zwei Confounding Events innerhalb der Event Windows festgestellt werden. Am 19. Dezember 2017 wurde

Tabelle 3: Gesamtmarktreaktion auf die Regulierung von PAs

	Ereignisdatum	Wahrscheinlichkeit der Regulierung	CR_{USA}	DJSTOXX Global 1800 ex America	$CMAR_{USA}$	$ACMAR_{USA}$	
(1)	20. Dezember 2017	+	0,0040	-0,0022	0,0062	0,0062	***
(2)	30. Juli 2018	+	-0,0093	0,0008	-0,0101	-0,0101	***
(3)	13. September 2018	+	0,0022	0,0139	-0,0117	-0,0117	***
(4)	8. November 2018	+	-0,0034	0,0037	-0,0071	-0,0071	***
(5)	15. November 2018	+	0,0030	-0,0036	0,0066	0,0066	***
(6)	21. August 2019	+	0,0011	0,0004	0,0008	0,0008	
(7)	31. Oktober 2019	-	0,0127	0,0115	0,0012	-0,0012	
(8)	5. November 2019	+	-0,0019	0,0015	-0,0034	-0,0034	***
(9)	30. Januar 2020	-	-0,0210	-0,0072	-0,0138	0,0138	***
(10)	22. Juli 2020	+	0,0082	0,0078	0,0004	0,0004	
(11)	18. September 2020	-	-0,0286	-0,0348	0,0063	-0,0063	***
	Mittlerer Return aller Ereignisse					-0,0010	
	t-statistic vs. 0					-5,5211	***
	Mittlerer Return der Non-Events					0,0010	
	t-statistic vs. Non-Events					8,7987	***

Anmerkungen: CR_{USA} ist der Cumulative Return der U.S.-Stichprobe über ein 3-Tages-Fenster, das um das Ereignis zentriert ist. DJSTOXX 1800 ex America ist der Vergleichsindex, der die 3-Tages-Cumulative Returns von 1200 Unternehmen aus der Asia-Pacific Region und Europa für jedes Ereignis umfasst. $CMAR_{USA}$ ist der Cumulative Market Adjusted Return (CR_{USA} - DJSTOXX Global 1800 ex America) für den U.S.-Markt. Wird erwartet, dass das Ereignis die Wahrscheinlichkeit der Einführung der Regulierung verringert (-), wird der $CMAR_{USA}$ mit -1 multipliziert ($ACMAR_{USA}$), um die Interpretation der aggregierten Ereignisse zu erleichtern. Genau umgekehrt wird allerdings die Reaktion des Marktes erwartet. Verringert ein Ereignis die Wahrscheinlichkeit der Einführung der Regulierung, wird demnach eine positive Reaktion des Marktes erwartet. t-statistic vs. 0 testet, ob sich der Mittelwert der 11 $ACMAR_{USA}$ signifikant von Null unterscheidet. t-statistic vs. Non-Events testet, ob sich die Stichprobe signifikant von Returns unterscheidet, die an Tagen ohne ein Ereignis aufgetreten sind. *, **, *** zeigen Signifikanz auf dem 10%-, 5%- bzw. 1%-Niveau an.

der Tax Cuts and Jobs Act im U.S.-Kongress von beiden Kammern beschlossen (Kaplan & Rappeport, 2017). Das neue Gesetz bedeutete deutliche Steuererleichterungen für U.S.-Unternehmen, die eine positive Marktreaktion hervorriefen. Es ist wahrscheinlich, dass der Tax Cuts and Jobs Act die Reaktionen auf den Corporate Governance Reform and Transparency Act vom 20. Dezember überdeckt hat. Das würde erklären, warum anstatt eines negativen, ein signifikant positiver CMAR beobachtet werden kann. Das zweite identifizierte Confounding Event betrifft den 30. Januar 2020. Anstatt einer erwarteten positiven Marktreaktion auf die Aussetzung des Gerichtsverfahrens und der Guidance durch die SEC, wird eine signifikant negative Reaktion beobachtet. Dies lässt sich wahrscheinlich auf den Ausbruch der Corona Pandemie zurückführen. Am 30. Januar 2020 verkündete die WHO in diesem Zusammenhang den internationalen Gesundheitsnotstand (Ghebreyesus, 2020). Um eine Verzerrung der Gesamtmarktreaktion durch die Confounding Events zu vermeiden, wurden die Ereignisse eins und neun ausgeschlossen und die Studie noch einmal durchgeführt. Dennoch zeigen beide t-Tests weiterhin eine signifikant negative Gesamtmarktreaktion auf dem 1%-Niveau (Anhang B).

Die Hypothese der negativen Gesamtmarktreaktion beruht auf der Annahme, dass sie vor allem auf Grund der Reaktionen institutioneller Investoren zustande kommt, für die die Regulierung Kosten darstellt und die gleichzeitig die größten Anteile an den börsennotierten Unternehmen halten (siehe Abschnitt 3.2.). In diesem Zusammenhang liegt die Folgerung nahe, dass die CMAR von Unternehmen mit einem hohen Anteil institutioneller Investoren eine stärkere, negative Marktreaktion zeigen. Um diese Folgerung zu überprüfen, wurde eine zusätzliche Analyse durchgeführt. Mit Hilfe der SEC Form 13F[6] Daten konnte den Unternehmen des finalen Datensatzes der jeweilige prozentuale Anteil institutioneller Investoren pro Ereignis zugeordnet werden. Anschließend wurden die Unternehmen in Quartile nach dem Anteil institutioneller Investoren geteilt. Ein t-Test untersucht für jedes Viertel, ob sich der Mittelwert seiner ACMAR signifikant von Null unterscheidet. Tabelle 4 fasst die Ergebnisse zusammen.

Man kann feststellen, dass die Marktreaktionen mit dem prozentualen Anteil institutioneller Investoren variieren. Der t-Test (t-statistic vs. 0) zeigt für jedes Viertel eine Signifikanz

[6] Form 13F ist ein Bericht, den institutionelle Investoren mit einem verwalteten Vermögen von mindestens 100 Millionen Dollar vierteljährlich an die SEC übermitteln müssen. Darin müssen sie unter anderem angeben, wie viele Aktien sie von welchem Unternehmen halten.

Tabelle 4: Gesamtmarktreaktion nach prozentualem Anteil institutioneller Investoren

Prozentualer Anteil instit. Investoren		$Q_{0,25}$ 16,62%	$Q_{0,50}$ 45,03%	$Q_{0,75}$ 75,91%
	1. Viertel	2. Viertel	3. Viertel	4. Viertel
Mittlerer Return aller Ereignisse	-0,0023	-0,0016	-0,0012	0,0009
t-statistic vs. 0	-4,5224 ***	-4,3756 ***	-3,9290 ***	3,3258 ***

Anmerkungen: Die Tabelle zeigt eine Aufteilung der Unternehmen nach dem prozentualen Anteil institutioneller Investoren in Quartile. t-statistic vs. 0 testet, ob sich der mittlere Return jedes Viertels signifikant von Null unterscheidet. *, **, *** zeigen Signifikanz auf dem 10%-, 5%- bzw. 1%-Niveau an. Das angegebene Signifikanzniveau gilt für jedes Viertel.

auf dem 1%-Niveau. Die Ergebnisse zeigen jedoch genau das Gegenteil der Annahme. Je höher der prozentuale Anteil institutioneller Investoren an den Unternehmen ist, desto geringer fällt die Marktreaktion auf die Regulierung von PAs aus. Das Viertel mit dem größten Anteil institutioneller Investoren zeigt sogar einen positiven mittleren Return (0,09%). Ein möglicher Grund dafür könnte in der Größe der Investoren liegen. Große institutionelle Investoren sind weniger auf PAs angewiesen, weil sie über ausreichend Ressourcen verfügen, um ihre Abstimmungsentscheidungen selbst zu erarbeiten (siehe Abschnitt 3.1.). Sie reagieren dementsprechend weniger auf die Regulierung von PAs. Zudem zeigt eine Studie der OECD, dass die drei größten institutionellen Investoren eines Unternehmens im Durchschnitt 25% der Anteile besitzen, die zehn größten besitzen bereits 45% und die 20 größten besitzen 60% (De La Cruz, Medina & Tang, 2019, S. 24). Daraus könnte man schließen, dass bei Unternehmen mit einem hohen Anteil institutioneller Investoren tendenziell mehr große Investoren vertreten sind. So ließen sich die geringer werdenden Reaktionen mit wachsendem Anteil erklären. Im Umkehrschluss ist es möglich, dass bei Unternehmen mit geringerem Anteil institutioneller Investoren gleichzeitig kleinere institutionelle Investoren vertreten sind, die von der Regulierung von PAs stärker betroffen sind, da sie über weniger Ressourcen für Analysen verfügen. Ihre Reaktion fällt dementsprechend stärker aus und könnte die der weniger aktiven Kleinanleger übertreffen, sodass es zu dem oben beobachteten Trend kommt. Auch diese Analyse wurde noch einmal ohne die CMAR der Confounding Events durchgeführt (Anhang C). Die Ergebnisse zeigen die gleiche Tendenz und sind ebenfalls signifikant auf dem 1%-Niveau.

4. Fazit

Um Aufschluss darüber zu geben, ob Aktionäre die Regulierung von PAs als Mehrwert oder als Nachteil auffassen, wurden in dieser Arbeit die U.S.-Marktreaktionen auf elf Ereignisse untersucht, die die Wahrscheinlichkeit der Einführung der Regulierung beeinflusst haben. Die Ereignisstudie zeigt eine negative Gesamtmarktreaktion, die darauf schließen lässt, dass Aktionäre die Regulierung weniger als Mehrwert durch erhöhte Transparenz, sondern vielmehr als Nachteil durch zusätzliche Kosten erwarten. Es wird festgestellt, dass die Reaktionen auf Ereignisse, die die Wahrscheinlichkeit der Einführung erhöhen (verringern), tendenziell negativ (positiv) ausfallen. Zudem wird festgestellt, dass die Marktreaktionen stärker ausfallen, je geringer der Anteil institutioneller Investoren an einem Unternehmen ist.

Die Erkenntnisse dieser Arbeit müssen jedoch mit Vorsicht interpretiert werden, da sie einigen Limitationen unterliegen. Zunächst ist es entscheidend, dass die richtigen Ereignisse identifiziert werden. Es kann nicht garantiert werden, dass die getroffene Auswahl vollständig die Ereignisse erfasst, die für die Reaktionen der Aktionäre ausschlaggebend waren. Zweitens kann die Aussagekraft durch Confounding Events verzerrt werden. Auch wenn nach dem Ausschluss der beiden identifizierten Störereignisse weiterhin eine signifikante Marktreaktion zu beobachten ist, kann auch hier nicht garantiert werden, dass ihre Identifikation vollständig ist. Drittens ist nicht auszuschließen, dass die Regulierung von PAs in den USA auch Auswirkungen auf den verwendeten Vergleichsindex hatte. ISS und GL sind global agierende Unternehmen. Ihre Regulierung könnte bei Aktionären in anderen Ländern zu der Befürchtung führen, dass das amerikanische Vorbild auch bei ihnen adaptiert wird. Zuletzt beziehen sich die Erkenntnisse dieser Arbeit auf die erwarteten Effekte der Regulierung, nicht auf die realisierten und sollten daher als vorläufige Evidenz angesehen werden (Joos & Leung, 2013).

Trotz der Limitationen trägt diese Arbeit zur Debatte über die Regulierung von PAs bei, indem sie als Erste empirische Erkenntnisse darüber liefert, wie der U.S.-Markt auf eine Regulierung reagiert. Außerdem zeigt sie weitere potenzielle Forschungsansätze auf. Die Untersuchung nach Besitzanteilen institutioneller Investoren weist darauf hin, dass die Marktreaktionen entlang bestimmter Unternehmenscharakteristika variieren können. Eine Querschnittsanalyse könnte weitere Erkenntnisse darüber liefern, welche Faktoren Einfluss auf die Marktreaktion haben. Es bleibt jedoch abzuwarten, wie die tatsächlichen Effekte nach Inkrafttreten der Regulierung ab Dezember 2021 ausfallen werden.

Literatur

2017/828/EC. (2017). *Directive (EU) 2017/828 of the European Parliament and of the Council of 17 May 2017 amending Directive 2007/36/EC as regards the encouragement of long-term shareholder engagement* (Report).

Alexander, C., Chen, M., Seppi, D. & Spatt, C. (2010). Interim News and the Role of Proxy Voting Advice. *Review of Financial Studies*, *23*, 4419–4454.

Balp, G. (2017). Regulating Proxy Advisors Through Transparency: Pros and Cons of the EU Approach. *European Company and Financial Law Review*, *14* (1), 1–36. Zugriff auf https://ssrn.com/abstract=3097721

Belinfanti, T. (2009). The Proxy Advisory & Corporate Governance Industry: The Case for Increased Oversight and Control. *Stanford Journal of Law Business & Finance*, *14* (2), 385. Zugriff auf https://ssrn.com/abstract=1558097

Brav, A., Jiang, W., Li, T. & Pinnington, J. (2019). Picking Friends Before Picking (Proxy) Fights: How Mutual Fund Voting Shapes Proxy Contests [Working Paper]. *European Corporate Governance Institute (ECGI)*.

Business Roundtable. (2020). *Public Comments to SEC on Amendments to Exemptions from the Proxy Rules for Proxy Voting Advice.* Zugriff auf https://www.businessroundtable.org/business-roundtable-public-comments-to-sec-on-amendments-to-exemptions-from-the-proxy-rules-for-proxy-voting-advice (Zuletzt aufgerufen am: 12.04.2021)

Cappucci, M. (2019). The Proxy War against Proxy Advisors. *Harvard Management Company*. Zugriff auf https://ssrn.com/abstract=3488427

Clayton, J. (2018). Statement Announcing SEC Staff Roundtable on the Proxy Process. *U.S. Securities and Exchange Commission*. Zugriff auf https://www.sec.gov/news/public-statement/statement-announcing-sec-staff-roundtable-proxy-process (Zuletzt aufgerufen am: 22.04.2021)

COEC. (2011). A call for change in the proxy advisory industry status quo: The case for greater accountability and oversight. (c11-07b).

Council of Institutional Investors (CII). (2020). *Comment letter to the SEC on proposed rules for proxy voting advice.* Zugriff auf https://www.sec.gov/comments/s7-22-19/s72219-6729687-207381.pdf (Zuletzt aufgerufen am: 20.04.2021)

De La Cruz, A., Medina, A. & Tang, Y. (2019). Owners of the World's Listed Companies. *OECD Capital Market Series*. Zugriff auf www.oecd.org/corporate/Owners-of-the-Worlds-Listed-Companies.htm

Department of Labor (DOL). (2020). *Fiduciary Duties Regarding Proxy Voting and Shareholder Rights, 81658-81695 29 CFR § 2509 and 2550.* Zugriff auf https://www.federalregister.gov/documents/2020/12/16/2020-27465/fiduciary-duties-regarding-proxy-voting-and-shareholder-rights

Department of Labor (DOL). (2021). *US Department of Labor Releases Statement on Enforcement of Its Final Rules on ESD Investments, Proxy Voting by Employee Benefit Plans.* Zugriff auf https://www.dol.gov/newsroom/releases/ebsa/ebsa20210310 (Zuletzt aufgerufen am: 20.04.2021)

Deutsche Börse. (2020). *Deutsche Börse übernimmt mit ISS führenden Anbieter von Governance-Lösungen, ESG-Daten und Analytik.* Zugriff auf https://www.deutsche-boerse.com/dbg-de/media/pressemitteilungen/Deutsche-B-rse-bernimmt-mit-ISS-f-hrenden-Anbieter-von-Governance-L-sungen-ESG-Daten-und-Analytik-2343864

Doyle, T. M. (2018). *The Realities of Robo-Voting* (Bericht). American Council for Capital Formation. Zugriff auf https://accfcorpgov.org/wp-content/uploads/ACCF-RoboVoting-Report_11_8_FINAL.pdf

Eckstein, A. (2015). Great Expectations: The Peril of an Expectations Gap in Proxy Advisory Firm Regulation. *Delaware Journal of Corporate Law*, *40* (77), 77. Zugriff auf https://ssrn.com/abstract=2469158

Ertimur, Y., Ferri, F. & Oesch, D. (2013). Shareholder Votes and Proxy Advisors: Evidence from Say on Pay. *Journal of Accounting Research*, *51* (5), 951–996.

European Securities and Markets Authority (ESMA). (2012). *An Overview of the Proxy Advisory Industry. Considerations on Possible Policy Options [Working Paper].* Zugriff auf https://www.esma.europa.eu/sites/default/files/library/2015/11/2012-212.pdf

Flook, B. (2020). *Institutional Shareholder Services Asks Judge to Toss Out Proxy Advice Rules, Guidance.* Zugriff auf https://tax.thomsonreuters.com/news/institutional-shareholder-services-asks-judge-to-toss-out-proxy-advice-rules-guidance/ (Zuletzt aufgerufen am: 20.04.2021)

Gallego Córcoles, A. (2016). Proxy Advisors in the Voting Process: Some Considerations for Future Regulation in Europe. *European Company and Financial Law Review*, *13* (1), 106–156. Zugriff auf https://ssrn.com/abstract=2842968

Ghebreyesus, T. A. (2020). WHO Director-General's statement on IHR Emergency Committee on Novel Coronavirus (2019-nCoV). *WHO*. Zugriff auf https://www.who.int/director-general/speeches/detail/who-director-general-s-statement-on-ihr-emergency-committee-on-novel-coronavirus-(2019-ncov)

Government Accountability Office (GAO). (2007). *Corporate Shareholder Meetings: Issues Relating to Firms That Advise Institutional Investors on Proxy Voting* (Report Nr. GAO-07-765).

Government Accountability Office (GAO). (2016). Corporate Shareholder Meetings: Proxy Advisory Firms' Role in Voting and Corporate Governance Practices, GAO-17–47. *U. S. G. P. Office*.

Griggs, N. (2019). More Than 300 Companies Join Nasdaq in SEC Letter. *NASDAQ*. Zugriff auf https://www.nasdaq.com/articles/more-300-companies-join-nasdaq-sec-letter-2019-02-05 (Zuletzt aufgerufen am: 10.04.2021)

Hayne, C. & Vance, M. (2019). Information Intermediary or De Facto Standard Setter? Field Evidence on the Indirect and Direct Influence of Proxy Advisors. *Journal of Accounting Research*, *57* (4), 969–1011.

Herren Lee, A. (2020). Paying More For Less: Higher Costs for Shareholders, Less Accountability for Management. *U.S. Securities and Exchange Commission*.

Hitz, J.-M. & Lehmann, N. (2018). Empirical Evidence on the Role of Proxy Advisors in European Capital Markets. *European Accounting Review*, *27* (4), 713–745.

Institutional Shareholder Services (ISS). (2019, October 31). *ISS Files Suit Over August SEC Guidance - Litigation to Prevent the Chill of Proxy Advisers' Protected Speech.* Zugriff auf https://www.issgovernance.com/iss-files-suit-over-august-sec-guidance/

Institutional Shareholder Services (ISS). (2021). *About ISS.* Zugriff auf https://www.issgovernance.com/about/about-iss/ (Zuletzt aufgerufen am: 10.04.2021)

Investor Advisory Commitee (IAC). (2020). *Recommendation of the SEC Investor Advisory Committee (IAC) Relating to SEC Guidance and Rule Proposals on Proxy Advisors and Shareholder Proposals.* U.S. Securities and Exchange Commission. Zugriff auf https://www.sec.gov/spotlight/investor-advisory-committee-2012/sec-guidance-and-rule-proposals-on-proxy-advisors-and-shareholder-proposals.pdf (Zuletzt aufgerufen am: 11.04.2021)

Joos, P. P. M. & Leung, E. (2013). Investor Perceptions of Potential IFRS Adoption in the United States. *The Accounting Review*, *88* (2), 577–609.

Kaplan, T. & Rappeport, A. (2017). Republican Tax Bill Passes Senate in 51-48 Vote. *The New York Times*.

Kiernan, P. (2020). SEC Votes to Regulate Proxy Advisers More Closely. *The Wall Street Journal*.

Klöhn, L. & Schwarz, P. (2013). The Regulation of Proxy Advisors. *Capital Markets Law Journal*, *8* (1), 90–107.

Larcker, D. F., McCall, A. L. & Ormazabal, G. (2015). Outsourcing Shareholder Voting to Proxy Advisory Firms. *The Journal of Law and Economics*, *58* (1), 173–204.

Larcker, D. F., Tayan, B. & Copland, J. R. (2018). The Big Thumb on the Scale: An Overview of the Proxy Advisory Industry. *Stanford Graduate School Research Paper*.

Malenko, N. & Shen, Y. (2016). The Role of Proxy Advisory Firms: Evidence from a Regression-Discontinuity Design. *Review of Financial Studies*, *29* (12), 3394–3427.

McElhaney, A. (2021). Ontario Teachers' and AIMCo Have Sold Glass Lewis to a Private Equity Firm. *Institutional Investor LLC*. Zugriff

auf https://www.institutionalinvestor.com/article/b1qztg15n9pf7s/Ontario-Teachers-and-AIMCo-Have-Sold-Glass-Lewis-to-a-Private-Equity-Firm (Zuletzt aufgerufen am: 10.04.2021)

Novick, B. & Cameron, R. (2018). November 2018 Roundtable Submission. *BlackRock, Inc*. Zugriff auf https://www.blackrock.com/corporate/literature/publication/sec-roundtable-proxy-process-111618.pdf (Zuletzt aufgerufen am: 20.04.2021)

Placenti, F. M. (2018). Are Proxy Advisors Really a Problem? *American Council for Capital Formation*. Zugriff auf https://accfcorpgov.org/wp-content/uploads/2018/10/ACCF_ProxyProblemReport_FINAL.pdf

Placenti, F. M. (2019). The SEC's Evolving Views Regarding Proxy Advisors. *Harvard Law School Forum on Corporate Governance*. Zugriff auf https://corpgov.law.harvard.edu/2019/11/23/the-secs-evolving-views-regarding-proxy-advisors/ (Zuletzt aufgerufen am: 06.04.2021)

Posner, C. (2018). SEC No-Action Letters Related to Proxy Advisory Firms. *Harvard Law School Forum on Corporate Governance*. Zugriff auf https://corpgov.law.harvard.edu/2018/09/15/sec-no-action-letters-related-to-proxy-advisory-firms/ (Zuletzt aufgerufen am: 11.04.2021)

Posner, C. (2020). SEC calls "time out" on proxy advisor guidance and ISS litigation. *Harvard Law School Forum on Corporate Governance*. Zugriff auf https://corpgov.law.harvard.edu/2020/02/14/sec-calls-time-out-on-proxy-advisor-guidance-and-iss-litigation/ (Zuletzt aufgerufen am: 20.04.2021)

ProxyPulse. (2020). 2020 Proxy Season Review. *PricewaterhouseCoopers LLP*. Zugriff auf https://www.broadridge.com/_assets/pdf/broadridge-proxypulse-2020-review.pdf

Retelny, G. (2019). Why we are suing the SEC. *Financial Times*.

Roisman, E. L. (2018). Statement at Proxy Process Roundtable. *U.S. Securities and Exchange Commission*. Zugriff auf https://www.sec.gov/news/public-statement/statement-roisman-111518 (Zuletzt aufgerufen am: 22.04.2021)

Roisman, E. L. (2020a). Myths and Realities: Modernizing the Proxy Rules. *U.S. Securities and Exchange Commission*. Zugriff auf https://www.sec.gov/news/speech/roisman-myths-and-realities-2020-01-30 (Zuletzt aufgerufen am: 10.04.2021)

Roisman, E. L. (2020b). *Open Meeting to Adopt Amendments to the Proxy Solicitation Rules.* Zugriff am 2021-04-20 auf https://www.sec.gov/news/public-statement/roisman-open-meeting-2020-07-22

Rose, P. & Walker, C. J. (2020). Examining the SEC's Proxy Advisor Rule. *The Ohio State University Moritz College of Law*.

Securities and Exchange Commission (SEC). (2010). Concept Release on the U.S. Proxy System. *Release No. 34-62495*. Zugriff auf https://www.sec.gov/rules/concept/2010/34-62495.pdf

Securities and Exchange Commission (SEC). (2013). SEC To Hold Roundtable On Proxy Advisory Services. *Press Release*. Zugriff auf https://www.sec.gov/news/press-release/2013-236

Securities and Exchange Commission (SEC). (2014). Proxy Voting: Proxy Voting Responsibilities of Investment Advisers and Availability of Exemptions from the Proxy Rules for Proxy Advisory Firms. *Staff Legal Bulletin No. 20*. Zugriff am 2021-04-11 auf https://www.sec.gov/rules/final/ia-2106.htm

Securities and Exchange Commission (SEC). (2018a). SEC Announces Agenda, Panelists for Staff Roundtable on the Proxy Process. *Press Release*. Zugriff auf https://www.sec.gov/news/press-release/2018-260

Securities and Exchange Commission (SEC). (2018b). Statement Regarding Staff Proxy Advisory Letters. Zugriff auf https://www.sec.gov/news/public-statement/statement-regarding-staff-proxy-advisory-letters (Zuletzt aufgerufen am: 12.04.2021)

Securities and Exchange Commission (SEC). (2019a). Amendments to Exemptions from the Proxy Rules for Proxy Voting Advice. *Release No. 34-87457*. Zugriff auf https://www.sec.gov/rules/proposed/2019/34-87457.pdf

Securities and Exchange Commission (SEC). (2019b). Commission Guidance Regarding Proxy Voting Responsibilities of Investment Advisers. *Release Nos. IA-5325; IC-33605*. Zugriff auf https://www.sec.gov/rules/interp/2019/ia-5325.pdf

Securities and Exchange Commission (SEC). (2019c). Commission Interpretation and Guidance Regarding the Applicability of the Proxy Rules. *Release No. 34-86721*. Zugriff auf https://www.sec.gov/rules/interp/2019/34-86721.pdf

Securities and Exchange Commission (SEC). (2020a). Exemptions from the Proxy Rules for Proxy Voting Advice. *Release No. 34-89372*. Zugriff auf https://www.sec.gov/rules/final/2020/34-89372.pdf

Securities and Exchange Commission (SEC). (2020b). SEC Adopts Rule Amendments to Provide Investors Using Proxy Voting Advice More Transparent, Accurate and Complete Information. *Press Release*. Zugriff auf https://www.sec.gov/news/press-release/2020-161

Securities and Exchange Commission (SEC). (2020c). Supplement to Commission Guidance Regarding Proxy Voting Responsibilities of Investment Advisers. *Release No. IA-5547*. Zugriff auf https://www.sec.gov/rules/policy/2020/ia-5547.pdf

Shubber, K. (2020). SEC approves new rules for proxy advisers. *Financial Times*. Zugriff auf https://www.ft.com/content/1d4fea3d-a250-4325-92fc-955c7444f383

Spatt, C. S. (2021). Proxy Advisory Firms, Governance, Market Failure, and Regulation. *The Review of Corporate Finance Studies*, *10* (1), 136–157.

Stein, K. M. (2018). Opening Remarks at the 2018 SEC Staff Roundtable on the Proxy Process. *U.S. Securities and Exchange Commission*. Zugriff auf https://www.sec.gov/news/public-statement/stein-remarks-2018-roundtable-proxy-process (Zuletzt aufgerufen am: 22.04.2021)

US-Congress. (2016). *Corporate Governance Reform and Transparency Act of 2016, H.R.5311, 114th Congress.* Zugriff auf https://www.congress.gov/bill/114th-congress/house-bill/5311/text

US-Congress. (2017). *Corporate Governance Reform and Transparency Act of 2017, H.R.4015, 115th Congress.* Zugriff auf https://www.congress.gov/115/bills/hr4015/BILLS-115hr4015rfs.pdf

US-Congress. (2018). *Corporate Governance Fairness Act, S.3614, 115th Congress* . Zugriff auf https://www.congress.gov/bill/115th-congress/senate-bill/3614/text?q=%7B%22search%22%3A%5B%22Corporate+Governance+Fairness+Act%22%5D%7D&r=1

US-Congress. (2019). *Corporate Governance Reform and Transparency Act of 2019, H.R.5116, 116th Congress.* Zugriff auf https://www.congress.gov/bill/116th-congress/house-bill/5116/text?r=80&s=1

Junior Management Science 8(1) (2023) 163-187

Junior Management Science

journal homepage: www.jums.academy

Memory asymmetries in experiential and material purchases: The role of self-expression

Anna Bogner

Technische Universität München

Abstract

To date many studies have investigated the different potentials of material and experiential purchases on making consumers happy. There is a strong recommendation to prefer experiential purchases when seeking happiness, the so called "experiential advantage". But do experiences always make people happier as compared to material purchases? This study proposes a memory asymmetry in experiential and material purchases as a main antecedent of reported consumer happiness in favor of experiential purchases. Employing a mixed online experiment, especially the role of self-expression in that context was investigated. Germanspeaking respondents indicated that when equally self-expressive, both material and experiential purchases can contribute to their happiness in the same degree. Nevertheless, an interesting finding was that there is a significant preference to use experiential purchases for conveying a person's true self to someone else. The heavier use of self-expressive purchases in socialization can explain why it is easier to recall self-expressive experiential than self-expressive material purchases. This memory asymmetry results in self-expressive experiential purchases being more frequently recalled than self-expressive material purchases when asked to intuitively recall a purchase that has contributed to one's happiness. Thus, that might be a reason why scholars have observed the "experiential advantage". The results are discussed as well as their meaning for different areas. Finally, recommendations for future research are offered.

Keywords: Material possessions; experiential advantage; memory asymmetry; happiness; self-expression.

1. Introduction

Happiness and the choices we make are strongly interrelated concepts and highly discussed topics nowadays. This thesis focuses on material vs. experiential purchases and their effect on consumer happiness. In our materialistic society, a great number of consumers think that money can buy happiness, but it has been shown that the wealth of society is not necessarily linked to increased happiness (Clark, 2017; Epstein, 1994). That is why many recommendations have been given to consumers with the aim to increase their happiness, such as using money to benefit others rather than themselves, buying many small pleasures rather than one big or delaying consumption or buying more experiences than material things (Dunn, Gilbert, & Wilson, 2011). Other studies have emphasized that consumers are made happier when their purchases are more central to their self-concept (Carter & Gilovich, 2012). Especially experiential purchases have been shown to have a self-expressive advantage (Bastos & Brucks, 2017; Kumar & Gilovich, 2015) and there is a strong recommendation to prefer experiential purchases when seeking happiness, the so called "experiential advantage", which has been demonstrated in many studies (Nicolao, Irwin, & Goodman, 2009; Peng & Ye, 2015; Van Boven & Gilovich, 2003). Nevertheless, as most of these studies rely on recalled memories of past purchases, it might be, that the experiential advantage is biased by memory. It has been shown that material purchases can be very relevant to the self and significantly contribute to one's happiness (Ahuvia, 2005; Belk, 1988; Caprariello & Reis, 2013; Guevarra & Howell, 2015). But when asked, it is the experiential purchases with a high conversational value and therefore higher potential for self-expression, that people also report to make them happier (Bastos, 2019; Kumar & Gilovich, 2015; Zhang, Howell, Caprariello, & Guevarra, 2014). Narratives can help consumers to make sense of personal experiences, justifying one's actions and maintaining a sense of self-worth (Baumeister & Newman, 1994). That is why the present thesis investigates whether there is an important effect of memory underlying the experiential advantage and whether the reported difference in happiness between expe-

DOI: https://doi.org/10.5282/jums/v8i1pp163-187

riential and material purchases can be better explained by what consumers remember and report instead of the purchase type itself. In other words, I propose that there exists a memory asymmetry with a bias towards recalling more self-expressive experiential purchases than self-expressive material purchases.

This study contributes to advancing research on consumer happiness and specially to shedding a new light on the experiential advantage. It aims to show the importance of considering human memory and memory biases when conducting studies that involve the recall and evaluation of past purchases on happiness, what hasn't been considered much in that area of research so far. It further wants to emphasize the importance of the self on consumer happiness, and specifically proposes self-expression as an important ingredient to happiness. Finally, it contributes to existing literature on the experience recommendation by showing that material and experiential purchases can make consumers equally happy when balanced on these aspects.

Firstly, investigating this theory is important for researchers because they might need to reconsider their recommendations. Most of them assume that happiness is the result of the distinction between experiential and material purchases. This study in contrast aims to show that there are other important factors involved, such as the level of self-expression. This would imply a shift of focus to investigating more specific characteristics of purchases. Secondly, it also affects policy makers because they might be urged to adapt their policies and recommendations to consumers concerning how to best spend their disposable income in order to increase their happiness in equal terms. That is why it also significantly matters to consumers. The results of this study could bring up a new aspect to the daily happiness discussion. They might further contribute to a shift from our experience society to some new and deeper values. Finally, it holds great importance for marketers. As people have increasingly more money available to spend on what makes them happy, it is important for marketers to know what that really is. The results of this study might give them important orientation for product design and marketing communication.

In continuation I give a ground theoretical background and revision of literature on happiness and subjective well-being, happiness in a consumer context, as well as self-expression as an antecedent. This is followed by the proposal of six hypotheses. The conducted experiment is described in detail together with a throughout analysis and description of the results. Following this, I discuss new insights drawn from the study, especially with regard on how they can contribute to current literature and why they matter to different stakeholders. I conclude by addressing limitations of the present research as well as giving recommendations for future research.

2. Theoretical Background

2.1. A Quest For Happiness

Happiness. The highest goal of humanity but as it seems, more and more difficult to achieve in our fast, overwhelming and materialistic society. How to find happiness?

As old as is humankind, as old is the search for the key to happiness. More often than not, it seems like a quest for the holy grail without clear indications where to search for. The great philosopher Aristotle showed this ambiguity and complexity of the concept when writing: "For this reason also the question is asked whether happiness is to be acquired by learning or by habituation or some other sort of training or comes in virtue of some divine providence or again by chance." (Ross, 1908, p. 1194). Ancient Greek philosophers such as Aristotle himself contributed to some of the earliest literature on happiness with a variety of different concepts on how to achieve it. Nevertheless, this prosperous time for philosophy and theory on happiness was replaced by centuries dominated by the Christian belief that happiness can only be found in the afterlife. Only when Michel de Montaigne and Baruch de Spinoza again took up ancient philosophical concepts in the eighteenth century, they initiated the modern quest for happiness (Lenoir, 2015). Still, theories on what causes happiness varied greatly. Whereas Jean-Jacques Rousseau saw its sources in a good bank account, a good cook and good digestion, his follower Henry David Thoreau, contributed it to activity (Diener, 2009). Although the sources of happiness had not been clearly defined yet, in the same century, it even became a human right when "the pursuit of happiness" was written into the American Declaration of Independence (Hood, 2019). The following century was not so bright for the happiness literature, as in the nineteenth century, within the romantic movement, happiness became highly critiqued and seen as a bourgeoise desire whereas unhappiness was perceived as being more authentic and creative. With the dramas and wars in the twentieth century, happiness even came to be regarded as utopic. This was clearly reflected in the work of philosophers of this century, like Martin Heidegger and Jean-Paul Sartre, focusing mainly on the question of anxiety. It was not before the 1960s when the concept of happiness gained interest again as effect of a synthesis of modern psychology and eastern spirituality. (Lenoir, 2015)

Today there are different theories concerning the sources of happiness. The telic theory for example proposes that happiness is gained when a goal is reached. This theory is strongly connected to the theory of pleasure and pain, because in order to formulate a goal, feeling a lack of something or deprivation is necessary (Diener, 2009). The activity theory assumes that happiness emerges from human activity such as described in the theory of flow by Csikszentmihalyi (1990) which proposes that activities are pleasurable when the level of challenge matches a person's skill level. The top-down theory sees happiness as the sum of many small pleasures whereas the bottom-up theory assumes that people in general tend to experience things positively, which

influences most momentary interactions of a person. Further, associationistic theories explore why some individuals have temperaments that seem more conducive to happiness. Judgment theories, in contrast, propose that happiness results from the comparison of a standard and the actual state. (Diener, 2009)

Such theories are subject of interest in a relatively new stream in psychology, called "positive psychology", which explores happiness and well-functioning individuals. It was first mentioned by Abraham Maslow and later introduced again by the former president of the American Psychological Association Martin Seligman in the 1990ies. In contrasts to the traditional psychology, centered on malfunctioning psychological processes, positive psychology deals with positive aspects of the human being such as happiness, optimism and character strengths. (Biswas-Diener & Dean, 2007)

"What is going right with people?" (Biswas-Diener & Dean, 2007, p. 11) is the central question in this stream of psychological research and results draw a clear picture of how a happy person looks like. Very happy people seem to have strong social relationships and spend little time alone (Diener & Seligman, 2002). In addition, they make more money, take fewer sick days from work, they have less work turnover, receive better supervisor evaluations on the job, get along with their colleagues better, are rated more highly by customers, spend more time volunteering and in contrast to less happy people are more likely to help strangers (Lyubomirsky, King, & Diener, 2005). Pavot and Diener (2008) researched life satisfaction judgments and showed that broad personality traits, such as extraversion and neuroticism play a significant role in how a person experiences his life. They discovered that these life satisfaction judgments are greatly influenced by individual differences as well as cultural norms and might change over time. In another study it was found that the level of life satisfaction and happiness tend to be adjusted after positive or negative life events. The same study discusses a number of things that seem to greatly influence people's happiness such as social ties, helping others or the opportunity to participate via democracy and federalism but also points out things that might be counterproductive for happiness. Those include extensive television viewing or social comparison (Bruni, 2009). The Austrian psychologist and holocaust survivor Victor Frankl, identifies happy people as those who have found their personal meaning (Frankl, 1985). This becomes especially clear in the preface of his book "Man's Search for Meaning" when he advises, "Don't aim at success—the more you aim at it and make it a target, the more you are going to miss it. For success, like happiness, cannot be pursued; it must ensue. . . as the unintended side-effect of one's personal dedication to a course greater than oneself." (Frankl, 1985, p. 12).

But what are we actually searching for? How does happiness feel like? Since Aristotle, scholars have distinguished between two types of happiness: hedonia, or feelings of moment-to-moment pleasure, and eudaimonia, or the broader sense of well-being that comes from the feeling that one's life is worthwhile, meaningful, and well-lived (Gilovich, Kumar, & Jampol, 2015a). Happiness is often thought of in the sense of hedonia, as an ecstatic feeling, a single ultimate goal. In contrast, research shows, that happiness is more a feeling of medium to moderately strong pleasant emotions most of the time (Diener & Seligman, 2002). According to Ed Diener, one of the pioneers in the field of subjective well-being, this kind of happiness contains both affective and cognitive components, one's situation and one's state of mind (Diener, 2009). Happiness is an interplay of a large number of emotional components, which differ from person to person. The concept of subjective well-being is therefore a global assessment of all aspects of a person's life (Diener, 2009). Lyubomirsky et al. (2005) proposed the theory that roughly half of our emotional well-being is due to genetics, that 10% is the result of life circumstances and that 40% is personal choices we make. The personal choices that we make every day are under our control, that is why this is the most relevant happiness component and one we can actively pursue. It sounds promising and simple but often becomes a great obstacle in our modern world. Which book to choose? The one with the compelling title "5 steps to happiness"? Or better "Reinvent your happiness"? Which one really contains the secret to happiness or is it just another promising marketing slogan that makes people buy yet another book that will later stand in their shelf, half read, dusty and forgotten?

2.2. Materialistic Consumption

The things we buy are often an intent to increase our happiness. During the last years the buying process has become easier and easier - one-click online purchase on Amazon, same-day delivery and free return - leading to impulsive buying and often not well-reflected choices.

Although scholars have tried to warn consumers of the negative aspects of our increasingly materialistic behavior, this tendency does not seem to stop accelerating. Some of these scholar's claims are that money cannot buy happiness and that pursuing materialistic ideals has a negative impact on happiness, letting unfulfilled basic psychological needs, lowering vitality and increasing depression and anxiety (Belk, 1985; Kasser, 2000; Shrum, Lowrey, & Pandelaere, 2014). Helga Dittmar and colleagues collected 750 measures from over 250 independent studies and demonstrated a clear and consistent negative association between personal well-being and people's belief in and prioritization of materialistic pursuits in life (Dittmar, Bond, Hurst, & Kasser, 2014). High-materialism consumers purchases are often motivated by showing extrinsic goals such as financial success, image or status (Kasser & Ryan, 1996). Belk (1985) points out that these materialistic and extrinsic oriented purchases might be the false sources of happiness to pursue. He further states, that sometimes people who have been disappointed in some way turn to materials to find their happiness there, but inevitably being disappointed again. Oftentimes, it is not the pleasure of acquisition but the pursuit that drives our material behavior (Hood, 2019), the need to keep consciousness

tuned (Csikszentmihalyi, 2000). In addition, some purchases can make us strive for more and more, leaving us with a feeling of restlessness and dissatisfaction. This is known as the "Diderot effect" which describes how individual items can have an effect on subsequent purchases. In 1969, the French philosopher Denis Diderot bought a new luxury dressing-gown but instead of feeling increased happiness as expected, it made him feel miserable. The new gown stood in strong contrast to his other possessions and he soon found himself buying new items of the same high quality. This increased spending and possessiveness made him feel bad and unhappy. (Hood, 2019)

In a study conducted by Hershfield, Mogilner, and Barnea (2016), the authors found, that materialistic behaviors are strongly enacted in today's society as most people favored more money instead of more time when they had the choice. Nevertheless, the one's choosing time over money were actually happier. Also, economists have observed a negative relationship between increasing wealth, materialism and happiness. The first to identify this effect was the economist Richard Easterlin in the 1970s, when he compared the increase in wealth of several economies after the second world war with the reported levels of happiness of their citizens (Easterlin, 1974). Since then, this has been known as the "Easterlin paradox". Clark (2017) also suggested a flat relationship between GDP per capita and average happiness in a country and proposed that this might be due to individuals tending to compare their income to a reference group. As their income rises, their reference group does shift too, leading to the same distance between their actual and desired state.

This clarifies, especially in a consumer context, that the choices we make are strongly connected with our subjective well-being and happiness. If more money and buying more things does not increase our happiness, how should our disposable income be spent?

2.3. The Experiential Advantage

A recent stream in literature suggests an answer to that question, proposing that we should shift from spending our money on material things to experiences. But when trying to differentiate and define experiential and material purchases, one quickly discovers that the distinction is not always clear-cut (Gilovich et al., 2015a; Van Boven & Gilovich, 2003). In fact, a number of studies have shown that some purchases can be framed in either experiential or material way, such as a BBQ grill which can be either assessed by its technical details or by the experience it provides with friends and family (Bastos, 2019; Bastos & Brucks, 2017). Nevertheless, some definitions have been made in literature to distinguish those purchase types, with the most recognized one by Van Boven and Gilovich (2003). In their work, they specify the following definition: "Experiential purchases are those made with the primary intention of acquiring a life experience: an event or series of events that one lives through. Material purchases are those made with the primary intention of acquiring a material good: a tangible object that is kept in one's possession." (Van Boven & Gilovich, 2003, p. 1194). According to this definition, the most relevant aspect for the differentiation of the two purchase types is their tangibility or intangibility.

The obvious differences between both purchase types and an increasing popularity of experiential purchases has led scholars to investigate the effects they have on consumer's happiness, assuming that there might be some differences. As early as in the 1970s, Alvin Toffler realized that consumers would one day "begin to collect experiences as consciously and passionately as they once collected things." (Toffler, 1970, p. 226). This obsession for experiences has come true and motivated more people than ever before to take one of the 47 Million flights worldwide in 2019 (IATA, 2020) to pursue an experience at a distant place. According to a Forbes article, 78% of millennials would rather spend their money on an experience than on a material good and living a less materialistic life, the so called "minimalism", has become increasingly popular among young generations (Weinswig, 2016). Nowadays we have reached a point where "a person's life is quite literally the sum of his or her experiences" (Van Boven & Gilovich, 2003, p. 1200). In line with this trend, it has been recommended by scholars to invest our disposal income in experiences (Gilovich, Kumar, & Jampol, 2015b; Howell & Hill, 2009; Nicolao et al., 2009; Van Boven & Gilovich, 2003). This shift of the happiness discussion in favor of "doing" instead of "having" is known as the "experiential advantage".

In several studies, Van Boven & Gilovich provided evidence for the assumption that experiential purchases make people happier than material purchases and first came up with the term experiential advantage. In detail, they found that experiences do not only make people happier and contribute to a better mood when thinking about them, but they are also seen as better financial investments. This was especially consistent among women, younger individuals and those living in urban or suburban communities (Van Boven & Gilovich, 2003). Those findings were successfully replicated and confirmed in studies by other scholars such as Howell and Hill (2009), Zhang et al. (2014) and Hayase and Ura (2015). Nicolao et al. (2009) came to similar conclusions but extending literature by showing that the effect is consistent in both directions. Experiences produce more happiness for positive purchases and less happiness for negative purchases than do material purchases. They also gave a potential explanation for the experiential advantage, indicating, that people adapt more quickly to their material purchases in contrast to their experiences (Dunn et al., 2011; Nicolao et al., 2009). Additionally, experiential purchases are in general more social, they include more people and therefore seem to provide higher levels of happiness than solitary purchases (Caprariello & Reis, 2013; Howell & Hill, 2009). They do not only better satisfy our psychological need for relatedness, they seem to be additionally of more hedonic nature which is positively related to well-being (Guevarra & Howell, 2015; Weingarten & Goodman, 2020). Another point is social comparison which has been found as an aspect counterproductive to happiness (Clark, 2017). In contrast to material things, experiences do not seem to be much subject to social com-

parison (Ang, Lim, & Leong, 2015; Carter & Gilovich, 2010; Howell & Hill, 2009). Moreover, people do not tend to rely as much on consumer reviews for experiential purchases than material purchases, indicating that they are less likely to reflect objective quality and are therefore more difficult to compare (Dai, Chan, & Mogilner, 2019; Dunn et al., 2011). As experiences are more unique and less interchangeable than material purchases, they additionally tend to generate less regrets of action but rather more regrets of inaction or missed opportunities (Rosenzweig & Gilovich, 2012). All these are arguments in favor of the experiential advantage.

It has been found that social class predicts purchase happiness derived from things vs. experiences. The experiential advantage emerges for higher levels of social class when all basic necessities and utilities are covered (Lee, Hall, & Wood, 2018; Weingarten & Goodman, 2020). In contrast, financially constrained consumers or consumers prioritizing a good use of one's money instead of happiness, seem to pursue material purchases that provide consumption utility over time (Pchelin & Howell, 2014; Tully, Hershfield, & Meyvis, 2015). Not only this group of consumers derives less happiness from their experiences, there are also those that consume their experiences with a conspicuous consumption intention in a way to simply "show off" (Duan & Dholakia, 2018). Especially today, many people tend to report all of their experiences on social media, always striving for more and more unique ones in order to meet the expectations of their audience, regardless of what makes them actually happy. In a study by Yu, Jing, Su, Zhou, and Nguyen (2016), another effect questioning the experiential advantage was found. High self-discrepancy consumers can be made happier by their material status purchases. Another weakness of experiences found in the happiness literature is, that people generally tend to derive greater happiness thinking about a past experiential purchase, but that their material purchases are the ones that made them happy more frequently (Weidman & Dunn, 2016). Further, material purchases that enable experiences make people equally happy than experiences (Guevarra & Howell, 2015). Finally, Caprariello and Reis (2013) found, that social purchases, independently of whether they are of experiential or material nature were rated as producing higher levels of happiness than solitary purchases.

To conclude, there is evidence for the assumption, that consumer happiness does not depend on the purchase type per se, but on other factors such as the specific properties of a purchase.

2.4. The Role of Self-Expression

As seen, happiness does not seem to be the sum of *how much* we can buy. Is it more about *what* we buy? Still, there are some studies that cast doubt that it is as simple as switching from material purchases to experiential ones and we become happier. Indeed it has been suggested that consumer happiness depends on whether our purchases are central to our self-concept (Carter & Gilovich, 2012). This theory lies in line with the argument that identity expression is related to increased happiness (Kifer, Heller, & Perunovic, 2013). All our purchases reflect our identity in some way and show who we are by the choices we make. Nevertheless, there can be differences in self-expression within experiential purchases, same which holds true for material purchases. My last trip to South America says a lot more about me than my last visit at the hairdresser. In the same sense, a flight ticket I purchased conveys better who I am than the hair dryer I bought. If self-expression is an important ingredient for happiness, it is again all about the right, deliberatively made purchase choices. But what are the purchases that express who I am? This points to another question which needs to be answered first: What is the self? And what is my true self or identity?

2.4.1. The self and expressing the self.

The question of the self leads us back to ancient philosophers. The famous sentence "I think, therefore I am" (cogito ergo sum) from René Descartes expresses his view of the self and its function very well. According to him it is a necessary condition in order to being able to believe in the existence of the external world. David Hume in contrast believed that the self is nothing more than the collection of different sensations and perceptions and John Locke added that it is our memory that gives us the feeling of continuity of our self. Immanuel Kant was the first one to come up with the theory that the self is the only thing of a human being that remains the same throughout time, independently of the experiences a person lives through. In contrast stands the so-called hermeneutical suggestion which sees the self as a narrative construction, something that evolves and changes according to one's projects and experiences and that is therefore bound to one's own interpretation. Finally, the phenomenological approach proposes that the self is an important condition for being able to experience reality, it is an integrated part of our conscious life. (Kircher & David, 2003)

What is sure is, that the self has been seen in different lights throughout history and its interpretation differs between fields of study (Kircher & David, 2003). Nowadays the self is mostly seen in the context of cultivation, in the sense of the hermeneutical suggestion. This implies that the self is a flexible concept and its construction is motivated by different goals of who we want to be as a person (Csikszentmihalyi & Halton, 1981). In this understanding, four studies conducted by Vignoles, Regalia, and Manzi (2006) aimed to test the influence of different motives on identity formation. They found that people were more likely to adapt identity elements that provided feelings of self-esteem, continuity, distinctiveness, belonging, efficacy and meaning. These findings were consistent on individual, relational and group level.

Additionally, it has been shown, that expressing those identity elements and one's self-concept in an authentic manner, such that they are consistent with the own physiological states, emotions, beliefs and cognitions, has a positive effect on a person's subjective well-being and is important for personal growth (Goldman & Kernis, 2002; Ryan, LaGuardia, & Rawsthorne, 2005; Wood, Linley, Maltby, Baliousis, & Joseph, 2008). According to H. S. Kim and Ko (2007) self-expression

can be defined as using ones words, choices or actions to express one's personality, traits, thoughts and feelings. Expressing oneself in such a way is further conductive to higher self-esteem (Lenton, Bruder, & Slabu, 2013) and can lower stress (Sheldon, Ryan, Rawsthorne, & Ilardi, 1997) whereas when restricted can promote inauthenticity (Kernis & Goldman, 2006), anxiety, low self-esteem and powerlessness (Wood et al., 2008). In conclusion, I propose the following hypothesis:

> *Hypothesis 1 (a): Self-expression is positively related to consumer happiness.*

Self-expression allows people to distinguish themselves from others and validate their concept of themselves. Belk (1985) argues in that way when saying that the identity of a person is an abstract concept without any concrete evidence, but that there are two primary ways for outward manifestation of the self: the things we have and the things we do.

2.4.2. Self-expressive material vs. experiential purchases.

Belk further states, that in our individualist culture, it is the choices we make which reflect our self where an object provides it with a body (Belk, 1988). In this context and according to his theory of the extended self, he describes objects as means to consciously expand our self to the exterior world. This aligns with the view of Sartre (2015) expressed in his work "Being and Nothingness" where he states that only by observing what we have we can know who we are. He suggests three ways through which an object can become to be regarded as part of ourselves. Through appropriating or controlling it, by creating it or by knowing it. Especially the use and display of products is essential for self-definition (Richins, 2004) and identity construction (Shrum et al., 2013) being strongest for high materialism consumers. Having a look around the homes of young adults, most often one can see future plans and goals reflected in their objects, whereas the homes of older married couples and objects in them tend to reflect shared past experiences (Belk, 1985).

These objects who have become part of ourselves and our identity can have a number of different functions. They can for example demarcate a boundary towards conflicting identities or help to support conflicting aspects of the self and provide a solution to an identity conflict (Ahuvia, 2005). In addition, material objects have the potential to connect one's meaningful past with the future through their permanence and as reminders of felt emotions (Goodman, Malkoc, & Stephenson, 2016).

As seen, material things can have a very important role in the identity construction of a person and significantly contribute to one's happiness through the symbolical meaning they provide and their potential to create order in consciousness. My favorite comfortable chair, the yoga mat on the floor, all the things in my home are connected to my sense of self, shape my awareness and order my consciousness. Although high-materialistic consumers are more likely to expect transformative properties of the objects they like to purchase (Richins, 2012), a study by Matz, Gladstone, and Stillwell (2016) found that generally, products that matched a consumer's personality provided high levels of satisfaction with life. Our favorite objects can therefore serve as individualized cues for self-expression (Wallendorf & Arnould, 1988). The objects we use stand in relation not only to ourselves but also to the people surrounding us and the whole universe. These three levels can be described by differentiation and integration. A person must first discover his own limits of being, then he must find ways to connect himself back to others (Csikszentmihalyi & Halton, 1981). The things we possess can express this dynamic process within people, they represent who we are as individuals and serve as symbols for our relationships with others (Wallendorf & Arnould, 1988). This holds especially true for things we feel strongly attached to and which we care for (Schultz, Kleine, & Kernan, 1989). An interesting and vivid example are special sub-cultures like Star-Wars movie fans. They might collect figures of characters from the movies, displaying them at a special place in their home. Those objects clearly express their self which is characterized by its passion for science-fiction, fantasy, creativity and the space, while at the same time indicating their belonging to the group of people which are "Star-Wars Fans".

As described, in contrast to a materialistic view on objects characterized by its possessiveness component, aimed to represent status and wealth, material purchases can also be seen in a symbolic view, so that they give meaning and order (Wallendorf & Arnould, 1988). Sometimes people even give up all their possessions with the aim to re-create themselves (Holt, 1997). This shows that material purchases have a great potential to express who we are, depending on how they are used and how wisely they were chosen. If chosen consciously, their great potential lies in their significance and ability to create a world of meanings that expresses our self, personal goals and thus has the potential to make us happy (Csikszentmihalyi & Halton, 1981).

Still, the happiness derived from objects or experiences changes throughout our life. Children generally derive more happiness from goods, but this changes in later childhood when they start to develop a greater sense for the people around them and their emotional intelligence increases (Chaplin, Lowrey, Ruvio, Shrum, & Vohs, 2020). Additionally, a study exploring the effect of experience types and age on happiness conducted by Bhattacharjee and Mogilner (2013), found self-definition as being the only driver of the effect. Whereas extraordinary experiences are self-defining at any age, ordinary experiences equally contribute to happiness as people age. This shows that self-expression is also an important factor contributing to the happiness we derive from our experiences. In fact, Gilovich et al. (2015b) argue that experiences are more self-expressive in general, because the same purchase seems to be less connected to the self when thought of in material terms rather than in experiential terms. Zhang et al. (2014) also address the finding that experiential buyers are happier with their life experiences than material buyers are with their material purchases, to increased identity expression. Even if materialists pursue material purchases with the intend to manage their identity, their experiences play an equally important part of that

identity (Carter & Gilovich, 2012). The concept of differentiation and integration which characterizes a well-functioning human being can be found in the context of experiences as well. It is true that experiences have the potential to provide the sense of an unique identity (Carter & Gilovich, 2012) while at the same time contributing to social connectedness through story-telling (Kumar & Gilovich, 2015). This can explain why people derive so much enjoyment and happiness from their experiences. People seem to be much aware if this fact. A study by J. Kim, Seto, Christy, and Hicks (2016) showed that people see experiences as a better investment for understanding one's true self and for exploring who they really are as a person than material purchases.

In the previous sections, we have seen that both material purchases and experiential purchases can be highly self-expressive. Further, simply consuming more is not related to subjective well-being and happiness, instead increasing the quality of what we consume might do its job (Csikszentmihalyi, 2000). This seems to be true for both purchase types. Choosing both more self-expressive things and experiences, purchases that really express who we are as a person, might increase the quality of our money spent significantly and as follows increase our happiness. That is why I propose the following hypothesis:

> *Hypothesis 1 (b): When purchases are equally self-expressive, purchase type (experiential vs. material) does not influence happiness.*

So finally what seems to really matter is not *how much* or *what* but *why* we choose something (Gilovich & Kumar, 2015). And choosing a purchase that is self-expressive, that expresses who we truly are as a person, might be a very good *why*.

2.5. Memory Asymmetries

2.5.1. Instrumentality for self-expression.

Although material and experiential purchases can be both self-expressive (Ahuvia, 2005; Carter & Gilovich, 2012) which is positively related to happiness, when asked, consumers tend to report that their experiential purchases made them happier (Gilovich et al., 2015b; Howell & Hill, 2009; Nicolao et al., 2009; Van Boven & Gilovich, 2003). Most of the studies supporting the experience recommendation are based on a memory recollection process of purchases made in the past. Given, that both purchase types can be self-expressive, it seems as though this recollection process might be distorted in favor of experiential purchases. In addition, people do not seem to be much aware of the advantage that material purchases have in some cases (Goodman et al., 2016).

Part of the explanation for this observation might be the higher instrumentality of experiential purchases to express ourselves. If something is a bigger part of ourselves, we are more likely to talk about it and connect with others through our reports of the purchases (Gilovich et al., 2015b). Social connection is an important ingredient to our happiness (Diener & Seligman, 2002). Especially in more individualist cultures, self-expression is a socializing function, driven by a self-verification process that makes the self recognizable and observable along with having positive psychological effects (H. S. Kim & Ko, 2007; Reed II, Forehand, Puntoni, & Warlop, 2012). In general there is no difference between the happiness derived from sharing about an experience or an object (Bastos, 2019). But although there might be both material and experiential purchases that highly express who we are as a person, people tend to talk more often about their experiential purchases in socialization. They seem to commonly belief that experiences can better convey who they really are. This has been shown by Kumar and Gilovich (2015), who found that people are in general more likely to choose an experiential purchase to talk about than a material purchase. This explains in part why people reported that experiences made them happier. Especially when telling their life story, when people really want to convey to someone else who they are, they prefer to use their experiential purchases (Carter & Gilovich, 2012). Bastos and Brucks (2017) additionally showed, that when framed in experiential terms, people attribute higher conversational value to a purchase, which was positively related to consumer happiness. They further argued that this effect was driven by the motivation for developing a relationship.

We tend to construct a life narrative, where we use possessions, events and places to infer the meaning of the self and describe our identity development (Schultz et al., 1989). Sometimes meaningful objects can help to narrate such stories but in general experiences are preferred, their "story value", as Gilovich et al. (2015a) describe it, is higher. Reason is that life narratives have a similar structure than experiences, including time sequences and relations of causality (Gallo, Townsend, & Alegre, 2019). In addition, experiential purchases are generally thought of in more intuitive and holistic terms (Gallo, Sood, & Mann, 2017; Gilovich & Gallo, 2020). This goes in line with the cognitive-experiential self-theory by Epstein (1994), in which he describes a memory system that is characterized by its holistic and affective nature. Past experiences are represented in images, metaphors and narratives. That is why narrating experiences requires less complex information processing (Epstein, 1994).

Experiences can support various interpretations and meanings, especially when communicating them to someone else (Baumeister & Newman, 1994). Our experiences only live on in our minds as mental representations of what we once lived through. These memories can easily be reinterpreted, altered or reworked (Dunn et al., 2011; Gilovich & Kumar, 2015; Marsh & Tversky, 2004). The satisfaction a material good provides is less memory-based, as it stays in one's possession for a longer time. It is not as easy to romanticize an old watch that stopped working or an uncomfortable chair as an experience we lived through some years ago (Gilovich et al., 2015a). This is related to the theory of the "rosy view" which suggests, that in retrospect, people tend to evaluate experiences much more positive than they actually were (Mitchell, Thompson, Peterson, & Cronk,

1997). All these are reasons for assuming that intuitively recalled purchases, as mostly observed in most previous studies exploring the experiential advantage, were not only of experiential nature but also more self-expressive. Therefore, I propose the following hypothesis:

Hypothesis 2 (a): Experiential (vs. material) purchases intuitively recalled from memory are more self-expressive.

2.5.2. Memory availability.

If experiential purchases are more instrumental for constructing our life narrative and convey it effectively to others in general, they will be consequently more often used in socialization. This leads to the assumption that because they are more often utilized, recall from memory is easier for more self-expressive experiential purchases than self-expressive material purchases.

That talking more about something increases its memorability goes back to basic functioning of human memory. It is divided in working memory and long-term memory. In order to improve the chances of making a good decision under time restriction and limited processing capacity, we tend to search for information first in working memory and use heuristics, simple clues, to do so (Payne, Bettman, & Johnson, 1991; Wyer & Srull, 1986). This implies that some information might be disregarded and overlooked but to be utilized, information must be readily available and processable (Payne et al., 1991). This is especially relevant for autobiographical memories, which are stored in the "episodic memory" (Tulving, 1972). Although it sounds so, this is not a fixed place in memory and therefore easy to locate. In contrast, information about the self is distributed throughout the whole memory system, there are bits of episodic memory everywhere and the information that one retrieves depends substantially on how the objects were processed at the time of access (Wyer & Srull, 1986).

A very effective way to increase availability of information in memory goes back to, according to Tversky and Kahneman (1973), one of the oldest laws of memory known. It says that repetition increases associative bonds. This has been shown in studies using monkey electrophysiology (Wirth et al., 2003) and human imaging (Law et al., 2005). Conclusions drawn from these experiments are, that the more often and excessive we have processed a given information, the higher the probability of spontaneously being able to retrieve it again from working memory (Habib & Nyberg, 2007; Wyer & Srull, 1986). Especially social rehearsal through storytelling, the process of sharing narratives with others, increases memorability and accessibility of information (Keven, 2016; Kulkofsky, Wang, & Ceci, 2008). That explains why, the more often we talk about a purchase, the more accessible it is in memory and the easier it is for us to retrieve the memory of this purchase. If people prefer to talk about their experiences to convey themselves to others, this might explain why, when recalling past purchases, they tend to recall more self-expressive experiential purchases than material purchases. As a consequence, they are rated as making people happier. Additionally, those self-expressive experiences might be perceived as being more central to the self and are again more readily remembered (Dunn et al., 2011). This leads me to proposing the following hypothesis:

Hypothesis 2 (b): It is easier to recall self-expressive experiential purchases than self-expressive material purchases.

It sometimes seems as if, although they are physically present, we forget our material purchases and it is our experiences that live on, through the stories that we tell. But maybe we do not forget our material purchases, we are rather not able to spontaneously access the information from long-term memory (Payne et al., 1991).

Combining the previous theoretical explanations, the aim of this study was to show that people remember and report more self-expressive experiences than self-expressive objects. As described, a potential reason for this memory asymmetry is that experiences are more instrumental for self-expression in general. People prefer to convey their self to someone else in the form of an experience rather than an object. Therefore, I propose the final two hypotheses, as mechanism underlying the previous effects:

Hypothesis 3 (a): Experiential purchases have an advantage in memory because they are more instrumental for communicating the self than material purchases.

Hypothesis 3 (b): Experiential purchases are therefore more often used in socialization.

Given the proposed relationship between self-expression and happiness, we can consequently no longer be sure whether it is the purchase type itself or rather the level of self-expression that generates consumer happiness. This would question the experience recommendation considerably.

3. Method

3.1. Research Design

To answer the hypotheses, I selected an experimental study employing a mixed-factor design. A mixed design is a research design that combines between-group and repeated-measures variables. It is an effective way to measure affective and cognitive processes.

Participants answered a survey with the title "Aspects of Consumer Behavior in the Context of Different Purchase Types". The survey platform Qualtrics was used to design and launch the survey. Before activating the survey, it was tested thoroughly. Convenient and snowball sampling was used to distribute it. Friends and family were asked to answer the survey and forward it to other people. In addition, it was shared on Social Media Survey Groups. Respondents were also recruited via the survey distribution platform SurveyCircle (www.surveycircle.com). It is a free online platform

where responses can be collected by supporting other people and participating in their studies. The amount of time needed to complete the survey was indicated to be approximately 10 minutes. The time frame of the study was 1 month from December 15, 2020 to January 14, 2021.

The experiment consisted of six different experimental conditions. After having read the instructions and given their consent to participate in the study, participants were randomly assigned to the between-subjects factor (low self-expression, high self-expression, control condition). Following this, they read a short explanation of self-expressiveness in purchases. Participants in the control condition did not read such a description. Employing a within-subjects factor, participants were then asked to recall a material or experiential purchase (in randomized order) they actually did within the last two years. To control for whether the instructions had been understood and induce people to think about the purchase, all participants were instructed to describe it. Following this, the dependent variables, ease of recall, happiness and instrumentality for self-expression were measured. In addition, the control variables, cost of the purchase, purchase nature (hedonic / utilitarian), purchase sociability, social comparison and purchase uniqueness were assessed. A manipulation check for self-expressiveness was additionally included in the questionnaire. Having completed these questions for the first purchase, participants were then asked to recall a second purchase (material or experiential depending on what they recalled before) and answer questions assessing the same set of dependent and control variables as before. Following this, a choice task was given aimed to assess instrumentality for self-expression in a social context when having both purchases in mind. Finally, before being thanked and debriefed, participants were asked for some demographic information. In the following sections, I will describe the procedure in more detail.

3.2. Procedure and Measures

3.2.1. Between-subjects factor: self-expression.

The aim of the study was to challenge the literature on the experiential advantage by showing that both purchase types (experiential vs. material) can make consumers equally happy when balanced on self-expressiveness. Further it intended to show that there exists an advantage of self-expressive experiential purchases in memory when spontaneously recalled. A reason for that can be the higher instrumentality of experiential purchases for communicating the self. In order to test the role of self-expressiveness in this context, it is therefore reasonable to use self-expression as between-subjects factor. According to this, three between-subjects conditions were part of the research design.

Participants were randomly assigned to one of three conditions, either a low self-expression condition, a high self-expression condition or a control condition. Drawing from Carter and Gilovich (2012) as well as Carroll and Ahuvia (2006), participants in the low and high self-expressive condition read the following description of self-expression in purchases (the control condition did not read such an introduction), before they were given additional instructions varying between the groups.

> *"Many of the purchases we make reflect our personality and express who we are to varying degrees. Sometimes, we make purchases that make a strong statement about who we are. Other times, our purchases, even though we might like them, reveal very little about our personality. The survey that follows is about a purchase you made in the past."*

Instructions for the high self-expression condition were:

> *"Think of an object (a tangible item) / experience (something intangible) that you bought during the last two years, that you feel strongly expresses your personality and shows who you truly are as a person. Please describe this object / experience."*

Participants in the low self-expression condition were told:

> *"Think of an object (a tangible item) / experience (something intangible) that you bought during the last two years, that you feel reveals very little about your personality and does not show who you truly are as a person. Please describe this object / experience."*

Finally, respondents in the control condition were given the following instruction:

> *"Think of an object (a tangible item) / experience (something intangible) that you bought during the last two years. Please describe this object / experience."*

Similar to Carter and Gilovich (2012) as well as Kumar and Gilovich (2015) a time frame was given in the instructions to control for temporal distance. The time frame of two years was used to allow for sufficient time so that people were probable to have made such a purchase while at the same time holding it as small as possible.

3.2.2. Within-subjects factor: experiential and material purchase.

In order to assess the effect of self-expressiveness in the context of different purchase types, the within-subjects part of the research design employed one factor with two levels, material and experiential purchases. Participants in all groups (low and high self-expression as well as control condition) were asked to think of a material and an experiential purchase (in randomized order) which they actually did. The instructions were consistent with Van Boven and Gilovich (2003) and Howell and Hill (2009). For the material purchase, they were given the following explanation and instruction:

> *"The survey that follows is about a purchase you made in the past. Specifically, it is about an object you purchased - a tangible item you bought. It was an object you could touch with your hand. Note that this purchase should not be one made to cover utilities and necessities."*

For the experiential purchase participants were instructed as follows:

> *"The survey that follows is about a purchase you made in the past. Specifically, it is about an experience you purchased - something intangible you bought. In other words, you did not end up with anything you could hold in your hand at the end of the experience. Note that this purchase should not be one made to cover utilities and necessities."*

3.2.3. Dependent measures and control variables.

After having read the instructions and described a material or experiential purchase, participants were asked several questions concerning the purchase mentioned. In the following presentation of the constructs and measurements employed, Cronbach's Alpha, in the case of multiple items (α), means (M) and standard deviation (SD) will be indicated.

Ease of Retrieval

The first dependent variable was ease of retrieval. After describing the recalled purchase, participants were asked "How difficult or easy was it to bring this specific type of purchase to your mind?" which they rated on a 7-point Likert scale ranging from 1 (very difficult) to 7 (very easy) (M = 5.08, SD = 2.08). In addition, the time between opening the question and writing the first character was measured in seconds, similar to a study conducted by Gardial and Biehal (1985) measuring memory accessibility and ease of retrieval, M = 27.12, SD = 52.20.

Cost of purchase

Following this, similar to Goodman et al. (2016); Richins (2012), participants were asked "How much did you approximately spend on this purchase (in €)?" and type the amount it in a text field, M = 1076.13, SD = 7112.41.

Happiness

A single item, adapted from Van Boven and Gilovich (2003), ("When you think about this purchase, how happy does it make you?") was used to measure the happiness a participant derived from his purchase. Answers were given on a 7-point Likert scale ranging from 1 (not happy) to 7 (very happy), M = 5.54, SD = 1.62.

Instrumentality for self-expression

Participants were then asked to answer two items indicating how much they disagreed or agreed with statements concerning the instrumentality of self-expression of the purchase on a 7-point Likert scale from 1 (totally disagree) to 7 (totally agree). The statements were "Telling other people about this purchase conveys well who I am as a person." and "Talking about this purchase helps to communicate myself to others.", 2 items, $\alpha = 0.92$, M = 3.76, SD = 1.90.

Purchase nature (utilitarian / hedonic)

Participants were also asked "Please rate whether you consider this purchase to be primary utilitarian (defined as useful, practical, functional, something that helps achieve a goal) or primary hedonic (defined as pleasant and fun, something that is enjoyable and appeals to the senses)", adapted from Gallo et al. (2017). They respondent on a 7-point semantic scale varying from 1 (mostly utilitarian) to 7 (mostly hedonic), M = 4.76, SD = 2.24.

Purchase sociability

Similarly, they read the question "To what extent did this purchase involve the presence of other people?" and indicated the sociability of their purchase on a 7-point Likert scale from 1(not at all) to 7 (very much), as in Caprariello and Reis (2013), M = 3.47, SD = 2.19.

Social comparison

Social comparison was assessed by adapting one item from Carter and Gilovich (2010), "Please imagine that someone else had made the same purchase that you had. How much would your satisfaction with your own purchase be diminished?" and rate it on a scale from 1 (not at all) to 7 (very much), M = 1.90, SD = 1.51.

Purchase uniqueness

Some studies also assess the uniqueness of the purchase (Bastos & Brucks, 2017; Hornik & Diesendruck, 2017). That is why participants were asked to respond to the question "How unique would you say this purchase was for you?" on a Likert scale from 1 (very little unique) to 7 (very much unique), M = 4.16, SD = 2.03.

Self-expressiveness

Finally, participants were asked to answer one item adapted from Carroll and Ahuvia (2006) serving as control item, concerning the self-expressiveness of the purchase. They were asked, "To what extend would you agree or disagree with the following statement about your purchase? - This purchase expresses who I am." and rate it on a Likert scale ranging from 1 (not at all) to 7 (very much), M = 3.64, SD = 2.00.

All these items were shown in this order two times, both for the experiential and for the material purchase.

Instrumentality for self-expression in socialization

In order to assess the instrumentality for self-expression imagining a real-life choice task and with both purchases mentioned previously in mind, participants were given the following indication: "Remember the two purchases you have described before (the material and the experiential one). Now imagine you are at a party. You don't know a lot of people and you start a conversation with someone you just met. If you only had to choose one of the two purchases, which one would you choose to better convey who you really are to this person?". They could then decide between choosing the material or the experiential purchase described before.

3.3. Sample

Three hundred fifty-nine responses to the online survey were collected within a time frame of one month. Participants who failed to correctly respond to the instructions, as assessed by visual revision of the purchase descriptions, were excluded from the data set. 43 participants did not correctly respond to the instructions for at least one purchase and were therefore excluded. In addition, seven responses were excluded due to missing data. The final sample included 309 responses.

Those respondents were distributed among the three experiential conditions as follows. There were 114 responses (36.9%) in the Control Condition, 101 responses (32.7%) in the High Self-Expressive Condition and 94 responses (30.4%) in the Low Self-Expressive Condition.

All respondents were able to describe their purchase in German, so I assumed that those participants did understand all instructions. It also means that the sample was completely German speaking and that all, or at least most of the respondents, were from Germany, Austria and Switzerland. The mean age of the sample was relatively young with 29 years (SD = 10.2, range = 17 – 78 years). Age did not differ much between the three groups. The sample was not evenly distributed among gender, there were more women (62%) than men (38%), but approximately this same proportion could be found in all three experiential groups. Only in the low self-expressive group the difference was smaller. The great majority of the participants had a Bachelor degree (43%), followed by 30% of participants having a High School degree and 14% a Master's degree. 7% had obtained a German Diplom. Only 4% achieved their Mittlere Reife and 2% left school after Hauptschule (comparable to middle school). Only a minority of participants archived their doctoral degree. Educational level of participants was relatively equally distributed among the three different experiential groups. In general, we can say that the sample was relatively high educated. The mean income of the sample was 2682€ (SD = 8023.98€ , range = 100 – 120000€). It can be seen that the standard deviation was very high. In addition, the mean income and standard deviation of the control condition (M_{CC} = 3324.69€ , SD_{CC} = 12056.50€) was higher than in the two other groups (M_{H-SE} = 2260.75€ , SD_{H-SE} = 3126.83€ ; M_{L-SE} = 2403.16€ , SD_{L-SE} = 5544.22€). This is due to some very extreme outliers. By excluding the three extreme outliers found (> 25000€), the mean income and standard deviation became smaller (M = 2130.81€ , SD = 2488.21€). Still, in the following analysis I did not exclude those outliers as I did not expect them to change the meaning of the outcome. In conclusion, the sample was relatively young with a majority of participants in the age of still attending university or having graduated recently.

3.4. Data Analysis

All statistical tests on the data were performed using the statistic program R. Mixed ANOVA was used to test the hypotheses 1(b) to 3(a). As Type III sums of squares were used, orthogonal contrasts were set for the predictor variables. To confirm the results from the ANOVA, a multilevel model was additionally built and tested for. The packages *ez* (for building the mixed ANOVA) *ggplot2* & *ggpubr* (for data visualization), *nlme* (for building a multilevel model), *pastecs* (for descriptive statistics) and *reshape* (for reshaping the data) were installed. The *rstatix* package for basic statistical tests was used for tests such as one-way ANOVA, Tukey HSD post-hoc test and related tests. It was also used for testing the dependent variables on homogeneity of variance using Leven's test of equal variances, Shapiro-Wilk test of normality and Box's M-test for testing homogeneity of covariances assumption. Chi-square tests for analysis with categorical variables as dependent and independent variables, especially employed in hypothesis 3(b), were performed using the 'CrossTable' function from the *gmodels* package. For regression analysis, as used to test hypothesis 1(a) and to build linear models, the 'lm' function, which comes within the *tidyverse* installation, was conducted. The *car* package was used for regression diagnostics. The package *corrr* and *corrplot* were downloaded to perform correlation analysis on the data and display a correlation matrix. Further packages used in the analysis were *readxl* for importing excel files, *textclean* for cleaning and processing text, *emmeans* for computing contrasts, *psych* for calculating cronbach's alpha and *knitr* to knit the rmarkdown document and export it as HTML file. The complete code and output can be found in the electronical appendix of this thesis.

4. Results

4.1. Manipulation Checks

The following analyses were intended to examine whether the manipulations employed in the online experiment were successful.

4.1.1. Group assignment and self-expression.

In order to test whether the between-subject manipulation of purchase self-expressiveness in form of a random assignment to one of three different experiential groups (control condition: CC / high self-expression: H-SE / low self-expression: L-SE) was successful, I compared the mean self-expressiveness of the purchases recalled in the three groups by means of a one-way ANOVA test. From this point on, experiential groups will always be used with their abbreviations.

Assumptions for conducting the ANOVA were not met, self-expression was significantly non-normal, $W = 0.98$, $p < 0.001$, this was also true among the groups. Homogeneity of variances could not be assumed as Levene's test was significant $p < 0.05$, which means that the variances in self-expression were different for study participants in all three groups, $F(2, 615) = 17.36$, $p < 0.05$ and the assumption was not tenable. Homogeneity of covariances was not given as well, as assessed by Box's test of equality of covariance matrices ($p < 0.001$). Nevertheless, as our sample size was big enough ($n > 30$), it was acceptable to violate the assumptions and compute the ANOVA test.

Table 1: Final Group Assignments

Group	*n*	%
Control Condition (CC)	114	36.9
High Self-Expression Condition (H-SE)	101	32.7
Low Self-Expression Condition (L-SE)	94	30.4

A one-way ANOVA was performed to evaluate if the purchase self-expression was different for the three experiential groups: CC (n = 114), H-SE (n = 101), L-SE (n = 94). Self-expression was statistically significant between different treatment groups, $F(2, 615) = 117.53$, $p < 0.01$, generalized eta squared = 0.28. Tukey HSD post-hoc test was performed on the data in order to see where the differences lied. Self-expression was significantly higher in the H-SE group compared to the CC, $b = 0.74$, $p < 0.001$. Self-expression was significantly lower in the L-SE group compared to the CC, b = -1.83, $p < 0.001$ and also significantly lower than in the H-SE group, b = -2.58, $p < 0.001$, ($M_{CC} = 3.95$, $M_{H-SE} = 4.69$, $M_{L-SE} = 2.12$). It can therefore be assumed that the experiential manipulation for self-expression was successful. Especially the difference in purchase self-expressiveness in the H-SE and L-SE was high. In contrast it can be seen already, that intuitively recalled purchases (CC), tend to range more at the high self-expressive end of the continuum and are therefore similar to purchases in the H-SE condition.

4.1.2. Purchase type.

The second manipulation check was aimed to test whether the within-subjects instructions concerning the type of purchase were successful. After participants were given the instructions according to their assigned groups, they were asked to write down and describe a material or experiential purchase they remembered (in randomized order). These descriptions were visually checked and decided whether participants had correctly understood the instructions and responded accordingly to them. Not meeting the instructions in this case meant exclusion from the data set. Responses were excluded in the case that (a) they did not describe the purchase type (experiential or material) that was requested in the instructions or (b) participants were not able to recall a purchase. Correct answers were coded with "Yes" (1), incorrect answers with "No" (2). Employing these criteria, overall, 11.98% of the initial 718 observations (counted by purchases not by participants) were excluded.

A chi-square test was used in order to determine whether there were statistically significant differences between the experiential conditions in terms of exclusion rates. Assumptions for conducting the test were met as 80% of expected cell counts were above 5 and no cells had an expected frequency below 1. The output from the chi-square test shows that there was significant evidence of a relationship between the groups and failure to correctly respond to the instructions, $t(2) = 12.87$, $p < 0.05$. In the control condition only 6.45% of assigned study participants failed to respond to the instructions. In the high self-expression condition 12.82% of responses were wrong. In the low self-expression condition 16.95% of respondents failed to respond in accordance to the instructions given. Although the exclusion rates differed across experiential conditions, the difference between groups was not too extreme. Hence, I concluded that excluding these responses from the data would not have an effect on the outcome of the following analysis. Nevertheless, this caveat needs to be considered when interpreting the results. In total, the instructions given concerning the purchase type to recall were met in 88.02% of cases. That is why I regarded them as being successful.

4.2. Answering the Hypotheses

4.2.1. Preliminary Analyses & Purchase Characteristics.

Before testing the single hypotheses, I tested basic associations between the depended, independent and control variables by means of a correlation table. Moderate correlations were found between the independent and dependent variables relevant for the hypotheses ($r > 0.3$, $r < 0.5$). The associations for the control variables were $r < 0.3$ in most cases, with exception of purchase nature (utilitarian / hedonic) and purchase uniqueness. The complete correlation table can be found in the R-markdown output in the appendix.

A summary of the purchase characteristics by purchase type (material vs. experiential) including means, standard deviation and main effect of the purchase type is presented in Table 2. Mixed ANOVA was performed with the purchase characteristics as dependent variables in order to see whether there were significant differences between the experiential conditions and purchase types. The variables were tested on normality assumptions by visual inspection of the data as well as performing the Shapiro-Wilk test. In addition, assumptions of homogeneity of variance were tested with the Levene's test as well as homogeneity of covariances assumptions by performing the Box's M-test. Although most of the assumptions were not met, it was still possible to perform the ANOVA on the data as the sample size was sufficiently big (n > 30).

The participant's purchases displayed a great variety in cost. Visual revision of the data showed that there were three very extreme outliers in the H-SE material condition. Those outliers were excluded for the following analyses of purchase cost but must be considered when interpreting it. Also, they were not excluded from the analysis considering the hypotheses, as it was assumed that they would not substantially change the results. Material purchases ranged from 1€ to 25 000€ (Mdn = 100€ , M = 495.40€) and

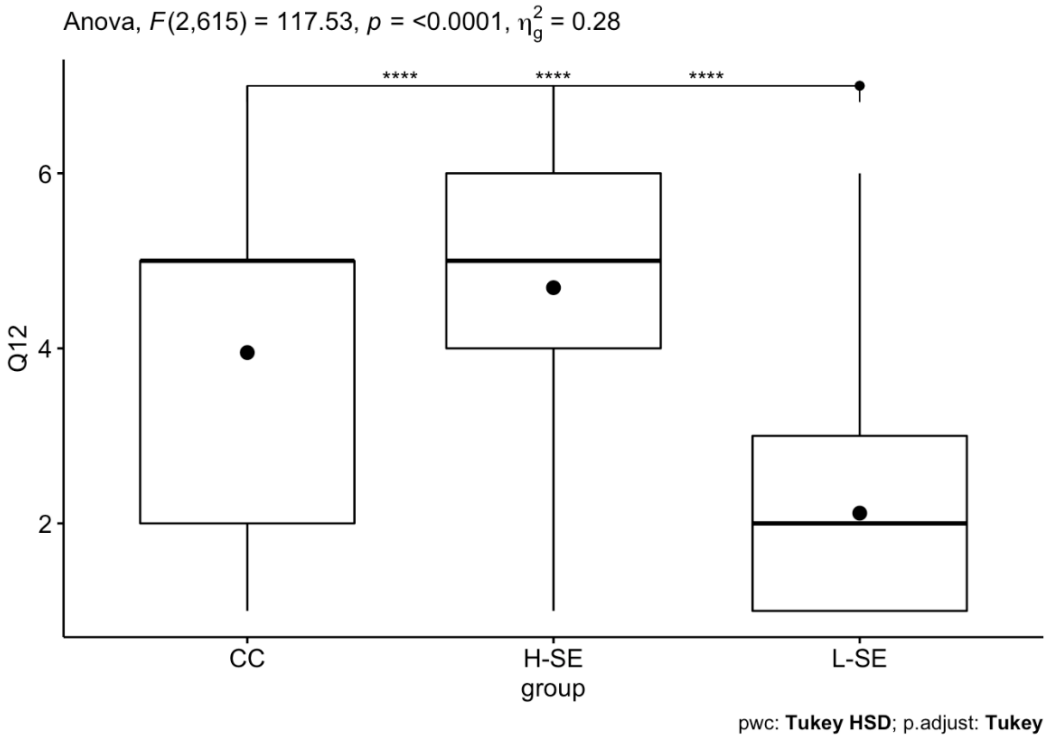

Figure 1: *Boxplot for Self-Expression by Group with Test Significances (ANOVA)*

experiential purchases ranged from 1€ to 20 000€ (Mdn = 100€ , M = 586.45€). Purchases in the CC ranged from 2€ to 20 000€ (Mdn = 149€ , M = 619.77€), purchases in the H-SE condition ranged from 2€ to 25 000€ (Mdn = 125€ , M = 663.81€) and in the L-SE condition from 1€ to 12 800€ (Mdn = 50, M = 317.18€). A two-way mixed ANOVA was performed in order to determine whether there were significant differences in cost between the experiential conditions, however no significant main or interaction effect was found. That is why I assumed that the mean purchase cost was similar in the experiential groups. Nevertheless, the three extreme outliers in the H-SE material condition need to be remembered, there might be, in some cases, a tendency to recall more expensive material purchases when they should represent one's true self.

Experiential purchases were found to be more hedonic and material purchases were more utilitarian in all three groups, however material purchases in the high self-expressive condition tended to be less utilitarian than purchases in the other conditions. Moreover, experiential purchases in the low self-expressive condition tended to be less hedonic than purchases in the other two conditions. This was confirmed by a significant main effect of the assignment group (and therefore self-expression) on the hedonic / utilitarian purchase nature $F(2, 306) = 12.83$, $p < 0.05$. The effect of purchase type on purchase nature was significant $F(1, 306) = 126.52$, $p < 0.05$. There was a significant interaction found between the group and purchase type, $F(2, 306) = 4.27$, $p < 0.05$. This indicates that the effect the assignment group had on the nature of a recalled purchase was different for experiential or material purchases. This was confirmed by a paired comparisons test.

It was also found that the sociability of experiential purchases was significantly higher (M = 4.27) than the sociability of material purchases (M = 2.67), independent of how self-expressive they were. It was deducted that a significant effect of purchase type existed on purchase sociability $F(1, 306) = 132.31$, $p < 0.05$. Pairwise comparison was performed to show that there indeed existed a statistically significant main effect of the purchase type on sociability, ($p < 0.001$).

Social comparison for material purchases was significantly higher (M = 2.08) than for experiential purchases (M = 1.73), independently of how self-expressive they were. The effect of purchase type on social comparison was significant $F(1, 306) = 16.40$, $p < 0.05$, pairwise comparisons confirmed this finding ($p < 0.01$).

Finally, there was evidence for a statistically significant main effect of the group on purchase uniqueness, $F(2, 306) = 18.57$, $p < 0.05$. Pairwise comparisons showed that purchase uniqueness was significantly lower in L-SE condition (m = 3.38) in comparison to the CC (m = 4.36), $p < 0.01$; as well as in comparison to the H-SE (m = 4.65), $p < 0.01$. In addition, it was also found to exist a statistically significant main effect of the purchase type on purchase uniqueness, $F(1, 306) = 28.32$, $p < 0.05$. Pairwise comparisons confirmed this finding ($p < 0.001$). This means that the ratings of purchase uniqueness for experiential purchases were

significantly higher (m = 4.52) than for material purchases (m = 3.79). No significant interaction effect was found. Given the differences in hedonic / utilitarian purchase nature, purchase sociability, social comparison and purchase uniqueness, those variables were used as covariates in the following hypothesis test.

4.2.2. Hypothesis 1(a).

In order to test whether self-expression was positively related to consumer happiness, a correlation analysis was executed. Spearman rho correlation coefficient was used because of the ordinal data employed. The correlation coefficient measures the level of the association between the two variables. In this case it was a strong positive effect ($r = 0.55$, $p < 0.05$).

Further it was of interest whether purchase self-expressiveness could significantly predict consumer happiness. That is why a regression analysis was conducted. A first linear regression model was built with self-expression as predictor for happiness. It was found that the self-expressiveness of a purchase significantly predicted the happiness derived from this purchase ($t = 15.94$, $p < 0.01$). The positive beta value ($b = 0.44$) indicated a positive relationship between self-expression and happiness. As self-expression of a purchase increased, the happiness derived from it also increased. The result of the regression indicated that the predictor explained 29.2% of the variance in happiness, as shown by R^2. An analysis of variance $F(1, 616) = 254.1$, $p < 0.01$ showed, that the model predicted happiness ratings significantly well.

A second model was tested to see whether the model was significantly improved when including the control variables that were found to vary between purchase types or experiential groups. Also, in this model, self-expressiveness was found to significantly predict happiness ($t = 12.50$, $p < 0.01$) in a positive direction ($b = 0.39$). Social comparison was found to predict purchase happiness ($t = -2.22$, $p < 0.01$) as well, but in a negative direction as the beta value showed ($b = -0.08$). This means, that when the ratings of a purchase were higher for social comparison, people were less happy with their purchase. Purchase uniqueness was also found to predict purchase happiness significantly ($t = 4.97$, $p < 0.01$) and in a positive direction ($b = 0.15$). The purchase nature (utilitarian/hedonic) and sociability did not significantly improve the model. The result of the regression in form of the R^2 value indicated, that the predictors explained 32.2% of the variance in happiness. Also this model predicted happiness significantly well, $F(5, 612) = 58.17$, $p < 0.01$. However, there was evidence that only the confident intervals of self-expression lied close together, indicating that those estimates were likely to be representative of the true population values. The interval for social comparison and purchase uniqueness were wider, indicating that those parameters were less representative, but nevertheless significant. The model was valid, there were no cases that had standardized residuals greater than 3, the histogram seemed relatively normally distributed, the scatterplot of standardized residuals did not show signs of heteroscedasticity, the VIF values were below 10 as well as the tolerance values above 0.2, hence there was no sign of multicollinearity.

Comparing the two models by means of an ANOVA, it could be seen that the second model significantly improved the fit of the model to the data compared to the first model, $F(4, 612) = 6.81$, $p < 0.001$. However, the F-value was relatively small. Also, the difference between the adjusted R^2 of the first model (adj. $R^2 = 0.29$) and the second model (adj. $R^2 = 0.32$) was not very big. Including the control variables into the model could only explain a 3.0% increase of variance in happiness of the population. That is why I assume that using self-expression for predicting happiness will do a quite good job.

In conclusion, Hypothesis 1(a) could be confirmed, consumer happiness is positively related to purchase self-expression and can above be predicted by it.

4.2.3. Hypothesis 1(b).

As the study confirmed that self-expressiveness can indeed predict purchase happiness, the next hypothesis aimed to test whether, when equally self-expressive, the purchase type (experiential vs. material) does not influence happiness. Mixed ANOVA was conducted as there were two independent nominal variables, one with independent groups and one with repeated measures. The data was plotted and purchase happiness was tested on normality assumptions by visual inspection of the data as well as by performing the Shapiro-Wilk test. In addition, assumptions of homogeneity of variance were tested with the Levene's test as well as homogeneity of covariances assumptions by performing the Box's M-test. Although the assumptions were not met, it was still possible to perform the ANOVA test on the data as the sample size was sufficiently big ($n > 30$).

All analyses for test assumptions were also performed preliminary to the hypotheses tests in the following sections employing mixed ANOVA. The depend variables in those tests did not meet the assumptions either, but due to the sufficiently big sample size, tests were carried out. This will be assumed for all the following analyses without mentioning it again.

It was found that a statistically significant mean effect of the experiential group existed on happiness, $F(2, 306) = 120.94$, $p < 0.01$. Pairwise comparisons showed that there was a significant effect between the CC and L-SE ($p < 0.001$) and between the H-SE and L-SE ($p < 0.001$). There was no significant difference between the CC and the H-SE group. The means of the CC ($M = 6.05$) and the H-SE ($M = 6.28$) were relatively similar, while the mean purchase happiness of the L-SE group ($M = 4.14$) was significantly lower. The effect of purchase type on happiness was not significant $F(1, 306) = 2.31$, $p > 0.05$ and a significant interaction effect was not found, $F(2, 306) = 1.86$, $p > 0.05$. As there are various advantages of analyzing repeated-measures data with a multilevel model, a GLM was conducted to further confirm the effect found in the two-way mixed ANOVA analysis. The model displayed the same main effect of the experiential group as

Table 2: *Purchase Characteristics: Means, Standard Deviations, and Main Effect of Purchase Type*

Measure	Material		Experiential			
	M	*SD*	*M*	*SD*	*F(2, 306)*	η^2
Purchase Cost	495	1933	586	1836	1.47	0.01
Utilitarian/Hedonic Purchase	3.61	2.12	5.91	1.70	12.83*	0.04
Purchase Sociability	2.67	1.91	4.27	2.10	0.40	0.00
Social Comparison	2.08	1.62	1.73	1.37	0.10	0.00
Purchase Uniqueness	3.79	2.01	4.53	2.00	18.57*	0.07

* p < .05. **p < .01. ***p < .001.

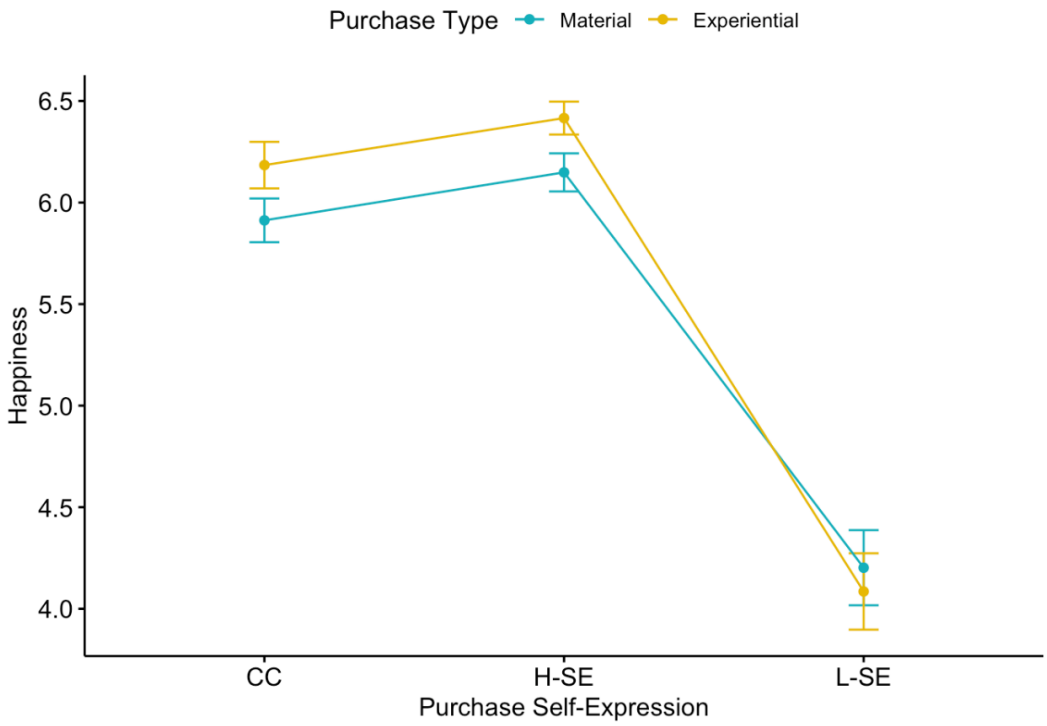

Figure 2: *Purchase Happiness by Group and Purchase Type*

found in the ANOVA test and no significant main effect for the type of purchase.

The non-significant main effect of purchase type and significant main effect of experiential group confirmed the hypothesis that when purchases are in the same self-expression group (high / low / control), being similar in self-expression, the type of purchase (experiential / material) has no effect on happiness. It was alone the difference in self-expressiveness between the groups that determined the happiness people derived from their purchases. Hypothesis 1(b) can therefore be confirmed.

4.2.4. Hypothesis 2(a).

Until now it has been shown that purchase self-expressiveness can predict purchase happiness and that purchases, independently whether they are material or experiential, can make people equally happy. Still, many studies have found, that when intuitively recalled, people tend to report more frequently experiential purchases in contrast to material purchases. That is why in hypothesis 2(a) I proposed, that this is because experiential (versus material) purchases intuitively recalled from memory are more self-expressive.

To test this hypothesis, a mixed ANOVA with self-expression as dependent variable was used to test whether significant differences in self-expression existed within the CC, where study participants intuitively recalled a purchase.

A significant main effect of the assignment group (and therefore self-expression) on happiness, $F(2, 306) = 95.73$, $p < 0.05$ was found. The effect of purchase type on happiness was also significant $F(1, 306) = 57.04$, $p < 0.05$, as well as there was a significant interaction found between the group and purchase type, $F(2, 306) = 8.25$, $p < 0.05$. This indicates that the effect the assignment group and therefore

Table 3: *Mixed ANOVA test output with Purchase Happiness as DV*

Predictor	df_{Num}	df_{Den}	SS_{Num}	SS_{Den}	F	p	η_g^2
(Intercept)	1	306	18518.10	679.01	8345.33	.000	.94
group	2	306	536.72	679.01	120.94	.000	.33
p.type	1	306	3.04	403.03	2.31	.130	.00
group x p.type	2	306	4.89	403.03	1.86	.158	.00

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. SS_{Num} indicates sum of squares numerator. SS_{Den} indicates sum of squares denominator. η_g^2 indicates generalized eta-squared.

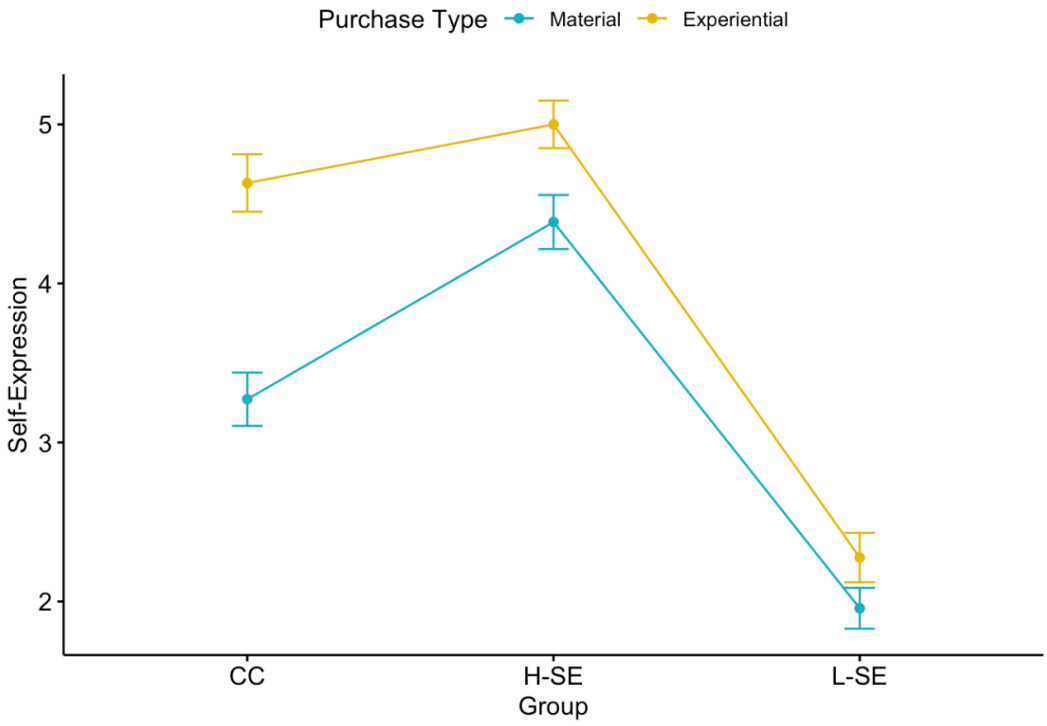

Figure 3: *Purchase Self-Expression by Group and Purchase Type*

self-expressiveness had on happiness was different for experiential and material purchases. By looking at the Bonferroni adjusted p-value (p.adj), it can be seen that the simple main effect of group was significant for experiential and material purchases in the case of the CC as well as the H-SE condition ($p < 0.001$). Pairwise comparisons show that the mean score for purchase self-expression was significant different in all group comparisons for material and experiential purchases ($p < 0.001$). The only comparison that was not significant was between the CC and H-SE condition for experiential purchases.

Due to the significant Bonferroni adjusted p-value in the case of the CC it can be said, that for intuitively recalled purchases, self-expression was significantly higher for experiential purchases ($M = 4.63$) than for material purchases ($M = 3.27$). An additionally performed GLM on the data, displayed the same effect. It confirmed that for the CC the difference in self-expression was significantly higher between the purchase groups than in the L-SE and H-SE condition ($p < 0.001$). We can therefore confirm hypothesis 2(a).

4.2.5. Hypothesis 2(b).

Experiential purchases intuitively recalled from memory were more self-expressive although it has been shown that both purchase types can actually be equally self-expressive and make people equally happy. Thus, the question why people tend to more frequently recall their self-expressive experiential purchases instead of their self-expressive material purchases still remains. In order to explain this effect, in hypothesis 2(b) I argued that it is easier to recall self-expressive experiential purchases than self-expressive material purchases from memory.

In order to test this hypothesis, two measures were employed. The first one recorded the time it took participants

Table 4: *Mixed ANOVA test output with Self-Expression as DV*

Predictor	df_{Num}	df_{Den}	SS_{Num}	SS_{Den}	F	p	η_g^2
(Intercept)	1	306	3560.53	1090.37	999.23	.000	.68
group	2	306	682.22	1090.37	95.73	.000	.29
p.type	1	306	105.37	565.31	57.04	.000	.06
group x p.type	2	306	30.47	565.31	8.25	.000	.02

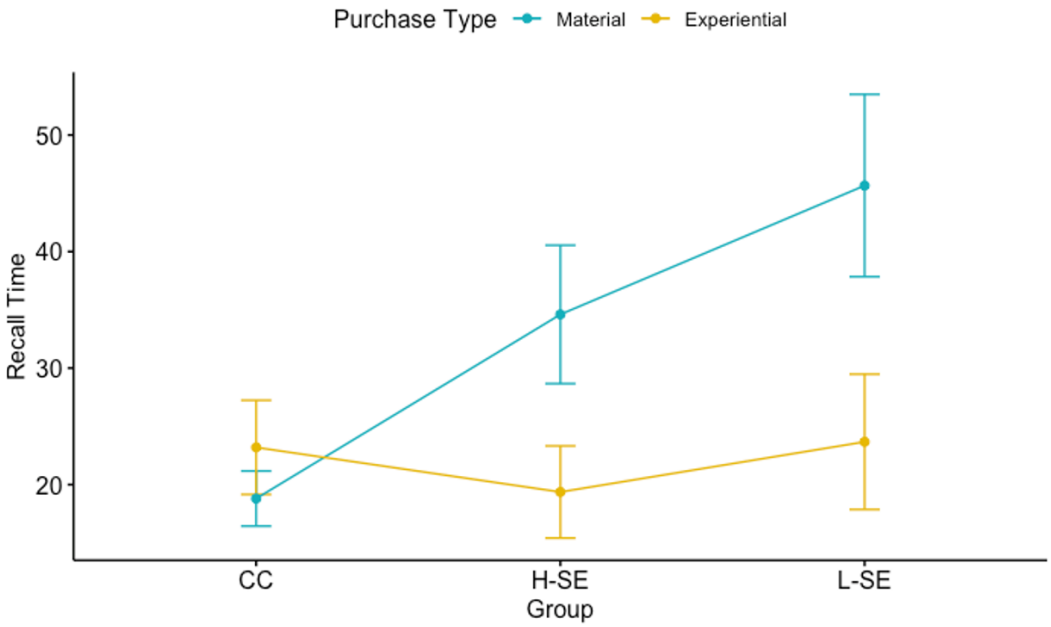

Figure 4: *Recall Time by Group and Purchase Type*

between opening the question asking for the description of a purchase they did, and responding to it. A mixed ANOVA with time of recall as dependent variable was used to test whether there were significant differences in recall time between the experiential conditions and the purchase types. A first look at the plotted data let one assume that there might indeed be an effect.

There was no significant main effect found of the purchase type on retrieval time, $F(1, 306) = 0.75$, $p > 0.05$. The effect of assignment group on retrieval time was not significant $F(2, 306) = 2.5$, $p > 0.05$ either. But there was a significant interaction found between the group and purchase type, $F(2, 306) = 6.83$, $p < 0.05$.

This indicates that the effect the assignment group and therefore self-expressiveness had on the recall time was different for experiential and material purchases. Considering the Bonferroni adjusted p-value (p.adj), it can be seen that the simple main effect of purchase type was significant only for the H-SE and L-SE condition ($p < 0.001$). Thus, while in the CC the time of recall was relatively similar for experiential and material purchases, it differed in the H-SE and L-SE group. Pairwise comparisons additionally showed that the mean score of recall time for experiential purchases was significantly different between the CC and H-SE group and the CC and the L-SE group ($p < 0.05$). This means that experiential purchases were relatively equally easy to recall between experiential groups, whereas it took participants longer to recall H-SE material purchases ($M = 34.61$) and L-SE material purchases ($M = 45.67$) in comparison to intuitively recalled material purchases ($M = 18.80$). Subsequently, performing a GLM on the data, revealed the same patterns. The significant main effect found for the H-SE condition, provided first evidence for the proposed hypothesis.

A second measure was used to test whether self-expressive experiential purchases were easier to recall than self-expressive material purchases. Also, in this case a mixed ANOVA test was employed to see whether participant's ratings of ease of recall differed between the experiential groups and purchase type.

There was a significant main effect found of the assignment group on ease of recall, $F(2, 306) = 54.03$, $p < 0.05$. The effect of purchase type on ease of recall was not significant $F(1, 306) = 0.12$, $p > 0.05$. There was no significant interaction between the group and purchase type, $F(2, 306) = 2.45$, $p > 0.05$. This indicates that the effect the assignment group and therefore self-expressiveness had on the ease of recall was not different for experiential and material purchases. Pairwise comparisons showed that there was a significant dif-

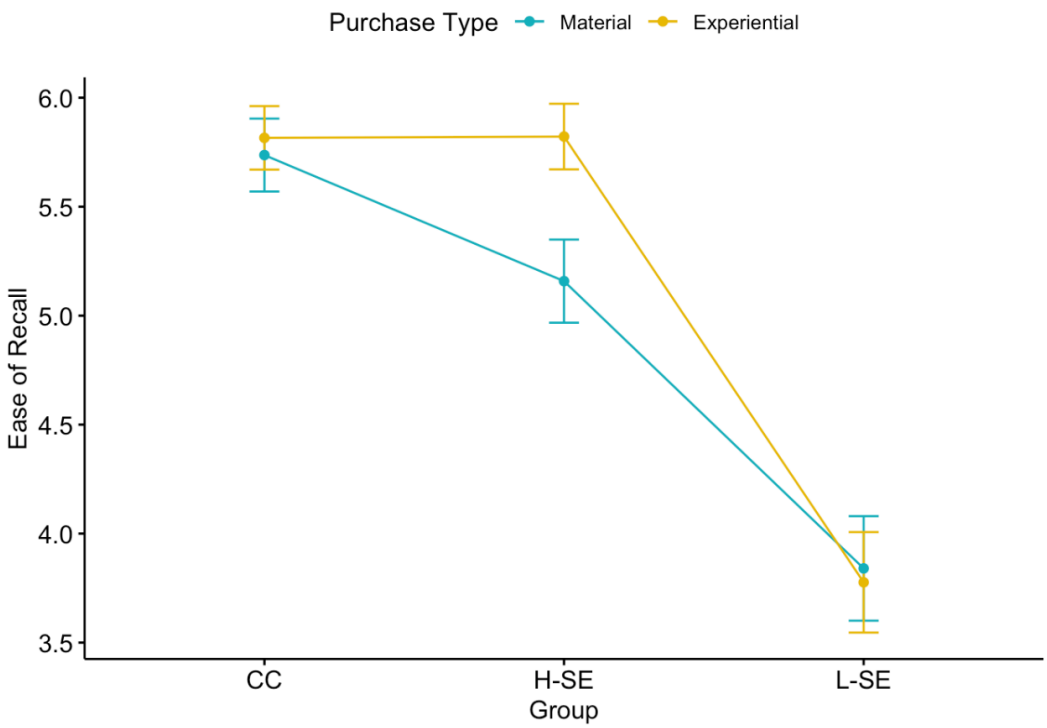

Figure 5: *Ease of Recall by Group and Purchase Type*

ference in ease of recall between the CC and L-SE condition ($p < 0.001$) as well as between the H-SE and L-SE condition ($p < 0.001$). In general, it was hardest to recall purchases in the L-SE condition ($M = 3.81$), followed by purchases in the H-SE condition ($M = 5.49$) and easiest for intuitively recalled purchases in the CC ($M = 5.78$). From the graphs it can be seen that it was easier to recall a self-expressive experiential purchase ($M = 5.82$) than a self-expressive material purchase ($M = 5.16$). Nevertheless, the effect was found to be insignificant. Performing a GLM showed similar results, with the exception that the difference in ease of recall between the CC and H-SE was found to be significant as well ($p < 0.01$). This slight difference was not of high importance, as it is not part of the hypothesis.

Although from the within-group comparisons of means, it could be seen that it was slightly easier to recall high self-expressive experiential purchases than material purchases, the effect was not found to be significant in the second test. This means that people did not necessary perceive their self-expressive experiential purchases to come easier to their mind as their self-expressive material purchases. Nevertheless, when looking at the time it took them to come up with those purchases, a significant difference could be observed, as found in the first part of this section. It took participants less time to recall self-expressive experiential purchases than self-expressive material purchases. Hypothesis 2(b) can therefore be confirmed as well.

4.2.6. Hypothesis 3(a).

Previously it has been shown that it was easier for study participants to recall self-expressive experiential purchases compared to self-expressive material purchases. This might be the reason why, when participants were asked for purchases that made them happy, intuitively recalled purchases tended to be more frequently of experiential nature and highly self-expressive. This was found in several other studies although we could confirm that both material and experiential purchases can make people equally happy. So why does this advantage of self-expressive purchases in memory exist? I proposed because experiential purchases are more instrumental for communicating the self to someone else than material purchases, as expressed in hypothesis 3(a).

Instrumentality for self-expression was measured employing two items. Both items displayed relatively similar patterns when looking at the descriptive statistics ($M_1 = 3.88$, $SD_1 = 1.97$; $M_2 = 3.67$, $SD_2 = 1.98$). In addition, the correlation coefficient between the two items (as measured by Spearman rho) lied close to 1 which indicated a strong positive correlation, rho $= 0.86$, $p < 0.001$. This means that as one item increased, the other one increased as well. As a measure of internal consistency or how closely related the two items were as a group, Cronbach's alpha was calculated, $\alpha = 0.96$. The result showed, that the scale was reliable. Therefore, it was possible to create a new variable for instrumentality for self-expression from the two items. In the following, two independent nominal variables, one with independent groups, one with repeated measures

Table 5: *Mixed ANOVA test output with Recall Time as DV*

Predictor	df_{Num}	df_{Den}	SS_{Num}	SS_{Den}	F	p	η_g^2
(Intercept)	1	306	100549.85	1179729.49	26.08	.000	.06
group	2	306	19256.26	1179729.49	2.50	.084	.01
p.type	1	306	1100.77	446904.85	0.75	.386	.00
group x p.type	2	306	19958.17	446904.85	6.83	.001	.01

Table 6: *Mixed ANOVA test output with Ease of Recall as DV*

Predictor	df_{Num}	df_{Den}	SS_{Num}	SS_{Den}	F	P	η_g^2
(Intercept)	1	306	7607.41	1269.18	1834.15	.000	.78
group	2	306	448.45	1269.18	54.06	.000	.17
p.type	1	306	0.36	927.23	0.12	.732	.00
group x p.type	2	306	14.84	927.23	2.45	.088	.01

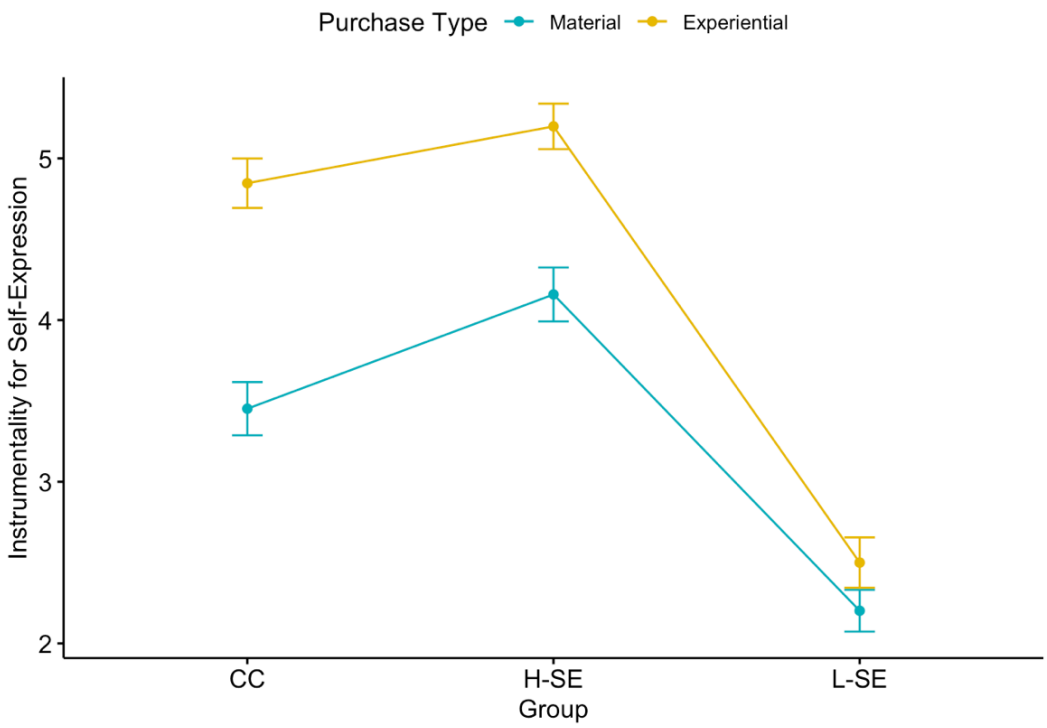

Figure 6: *Instrumentality for Self-Expression by Group and Purchase Type*

Table 7: *Mixed ANOVA test output with Instrumentality for Self-Expression as DV*

Predictor	df_{Num}	df_{Den}	SS_{Num}	SS_{Den}	F	p	η_g^2
(Intercept)	1	306	3925.07	1043.84	1150.62	.000	.72
group	2	306	577.89	1043.84	84.70	.000	.28
p.type	1	306	110.88	446.87	75.93	.000	.07
group x p.type	2	306	31.66	446.87	10.84	.000	.02

were used to conduct a mixed ANOVA on instrumentality for self-expression. Before, descriptive statistics as well as test assumptions were performed and the data graphically displayed. The test output showed a significant main effect of the experiential group on instrumentality for self-expression, $F(1, 306) = 84.70$, $p < 0.05$. The effect of purchase type

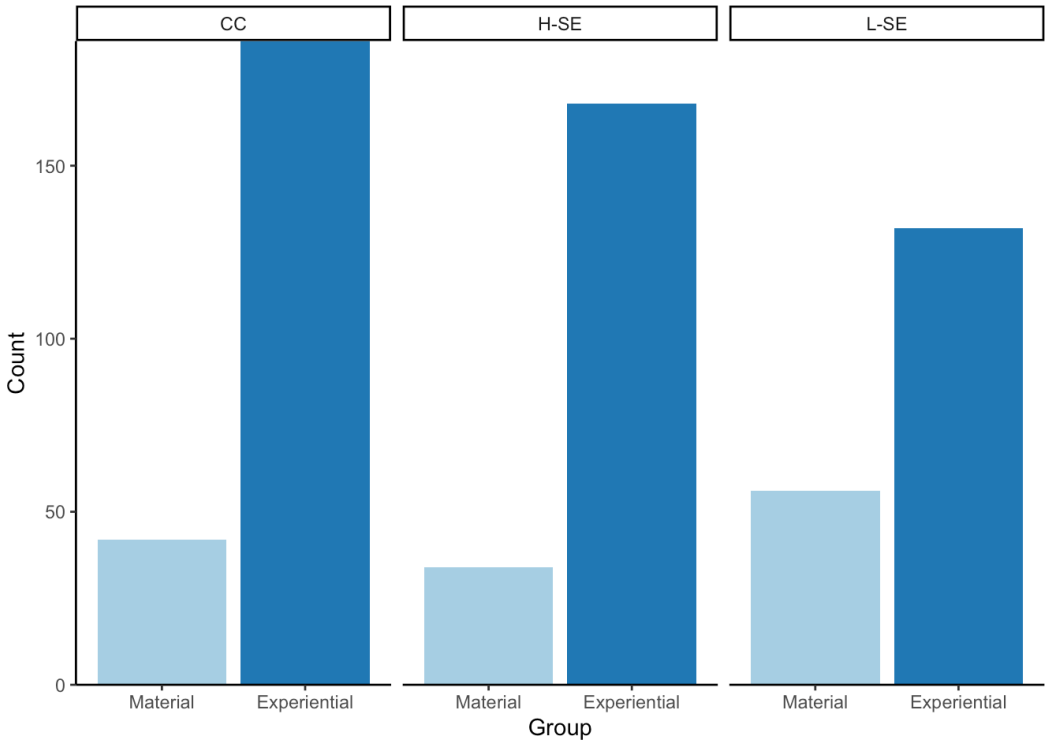

Figure 7: *Instrumentality for Self-Expression in a Social Context by Group and Purchase Type*

on instrumentality for self-expression was significant as well, $F(2, 306) = 75.93$, $p < 0.05$. It was found that a significant interaction existed between the group and purchase type, $F(2, 306) = 10.84$, $p < 0.05$. Considering the Bonferroni adjusted p-value (p.adj), it can be seen that the simple main effect of purchase type was significant for the CC and H-SE condition ($p < 0.001$). Pairwise comparisons confirmed that the mean score for instrumentality for self-expression in the CC was significantly higher for experiential ($M = 4.85$) than material purchases ($M = 3.45$), $p < 0.001$. Also, in the H-SE condition experiential purchases ($M = 4.16$) were more instrumental for self-expression than material purchases ($M = 5.20$), $p < 0.001$. Performing a GLM on the data displayed the same effects.

The findings confirmed hypothesis 3(a) and showed clearly that experiential purchases were more instrumental for self-expression, although both experiential and material purchases can express well one's identity. This held true for all three between-subjects groups and therefore independently of whether the purchase was self-expressive or not, but the effect was only significant for the intuitively recalled purchases and high self-expressive purchases. Nevertheless, the ambiguity of the question for low self-expressive purchases could be the reason for this finding.

4.2.7. Hypothesis 3(b).

An advantage in memory does not exist simply because of the preference for something, people would need to actively talk more often about their experiential purchases for this effect to emerge. Due to that, hypothesis 3(b) intended to test whether people indeed use their experiential purchases more often in socialization and what would explain this asymmetry in memory. In a choice task, people were asked to imagine a social situation and deliberately choose between expressing their true self to someone else by talking about their material or their experiential purchase they had mentioned before. As this hypothesis dealt with nominal variables, a chi-square test was used in order to determine whether there was a significant difference in how often the purchase types (experiential vs. material) were used in the choice task simulating a social situation. Assumptions for performing the test were met as 80% of expected cell counts were above 5 and no cells had an expected frequency below 1. The results showed that in a great majority of the cases, people chose their experiential purchase to convey their true self to someone in the situation (78.64%). This held true for the CC (81.58%) as well as for the H-SE condition (83.17%) and L-SE condition (70.21%). Still, a significant evidence of a relationship was found between the groups and which type of purchase was preferred for self-expression, $t(2) = 11.59$, $p < 0.05$. In the L-SE group, there was a tendency to more frequently choose the material purchase for expressing the self (29.79%) than in the other conditions. The reason could be the ambiguity of the question in that case. Nevertheless, participants preferred to talk about their experiences in all three groups.

That is why also hypothesis 3(b) can be confirmed, experiential purchases were more often used in socialization and this held true for high self-expressive purchases as well as low

self-expressive purchases and intuitively recalled purchases. It can therefore be said that in socialization, people prefer to convey their self to another person by means of the experiences they purchase. That is why those purchases are heavily used and gain an advantage in memory over time.

5. Discussion

Consumers often purchase experiences and goods to increase their happiness (Richins, 2012). To date many studies have investigated the different potentials of material and experiential purchases on making consumers happy and there is a strong recommendation to prefer experiential purchases when seeking happiness, the so called "experiential advantage", which has been demonstrated in many studies (Gilovich et al., 2015b; Howell & Hill, 2009; Nicolao et al., 2009; Van Boven & Gilovich, 2003). The aim of this study was to advance research on consumer happiness and especially shed a new light on the experiential advantage. The study does so and contributes to the literature by showing the importance of considering human memory effects when conducting studies that involve the recall and evaluation of past purchases on happiness. In line with previous findings by Carter and Gilovich (2012), it has been shown, that people generally prefer to use their experiential purchases to convey who they are as a person to someone else. They also tend to actively talk more often about their experiences in socialization. The relevance of experiential purchases to the self and motivation to talk about them can be exemplary seen by the following description of a study participant:

> *"A festival. This may not be particularly unique for me, but festivals etc. play a big role in my life, as they reflect certain joie de vivre and serenity alongside the music, which I want to convey."*

The heavier use of such self-expressive experiential purchases in socialization leads to their advantage in memory. This was demonstrated by employing a control condition in which study participants were asked to intuitively recall a purchase. More self-expressive experiential than material purchases were recalled. Also, participants reported it would be easier to recall self-expressive experiential than material purchases. As it was shown that purchase self-expression can predict purchase happiness, it is now reasonable that this memory asymmetry could be a reason for why the experiential advantage has been found and replicated in so many studies. The present study advances consumer literature by demonstrating the effects of memory behind the experiential advantage and by emphasizing the relevance of self-expression for consumer happiness. It is a main ingredient that can significantly contribute to the happiness we derive from a purchase. Considering this, material purchases can contribute to consumer's happiness in the same degree as experiential purchases. The following example from the study vividly shows this observation:

> *"The item is white boots with a lace-up and a thick white sole with a striking tread. Why I chose this item is probably because I like to express my personality with clothes. I myself am relatively small, I have blonde hair and a rather doll-like face. I love the boots so much because they allow me to show that I am not so "smooth" but "edgy"."*

Also Goldman and Kernis (2002) support this view of self-expression as main antecedent of consumer happiness by arguing, that rather than shifting spending from material to experiential purchases, integrating material and experiential consumption is key to a happy life in a consumerist society. The findings of this study are not only important to researchers but also to marketers, consumers and policy makers.

They are important to marketers because they could consider these aspects when designing and promoting their products. The relevance is two-fold, considering the increased happiness consumers derive from purchases that express who they are as a person and the great potential of experiential purchases being more often talked about. It has been shown previously that enhancing self-expressive aspects of an offering has positive effects on brand love (Carroll & Ahuvia, 2006). This means that offerings that achieve to express the self of a person might also increase their emotional response towards the brand in a positive way. This is important because it has been found that such positive feeling towards the product or brand relate directly to product satisfaction judgments, complaint behavior and volume of word-of-mouth transmission (Carroll & Ahuvia, 2006; Westbrook, 1987). Increasing the potential for self-expression of a purchase, might lead to higher customer loyalty and more frequent WOM. A potential way to increase self-expression would be to target small and individualized customer segments and adapt the characteristics of the object or experience to the specific needs of that target group. Another way could be customization or co-creation of products.

Marketers can also learn from the advantage of experiences for expressing one's true self in a social context for their material offerings. Framing them in a more narrative and symbolic way could enhance their potential of being retrospectively remembered and more often talked about, hence increasing desirability and word of mouth. In fact, Gallo et al. (2019) recently found that when products are framed in an experiential way, people are more likely to review it. Communicating offerings in such a way is known as experiential marketing, where it is not the resource that is in the center of importance but the perspectives and meanings that are presented to the consumer (Lindgren, Vanhamme, & Beverland, 2009). Symbols are key in this marketing approach, that is why it is very much in accordance with the importance of purchase self-expression found in this study. Lindgren et al. (2009) suggest to marketers to define in detail which perspectives and meanings they want to convey with their offering, in order to achieve a connection between the product and consumer. They further emphasize, that a mem-

orable customer experience integrates both emotional and rational aspects in a valuable combination. One practical example would be to invite consumers to a show where the products are used or connect them to special life events, like one's graduation or wedding (Goodman et al., 2016). Further, marketers could use backward framing via post-contact e-mailing and advertising to keep the memory and meaning of the purchase alive (Lindgren et al., 2009).

For consumers, the results of this study could be a motivation to be more conscious when making a purchase. The study emphasizes the idea that it is not as simple as spending one's disposable income on less material things and more experiences. Rather consumers could understand that investing in purchases that help to understand who they are as a person and maximize their individual self-expression can have a much bigger impact on their happiness and well-being. This is aligned with Matz et al. (2016) who suggested that there are optimal and suboptimal ways of spending, depending on personality. Suboptimal spending is very nicely displayed in the following purchase descriptions from the survey:

> *"It was a ticket to an amusement park that I thought would be a bit calmer, in reality all the options were either too extreme or too childish for me, I would have been happy to save that money"*

Or in the case of a material purchase:

> *"It was a prayer rug, really nice, and actually I wanted it as an inspiration to find more connection to faith. To be honest I just came to the realization that I am not a religious person and the thing is now gathering dust in my dresser drawer".*

Taking the results of this study together, consumers could learn to better evaluate their different purchase possibilities on their potential to increase happiness and help them considerably to make better choices. In addition, focusing more on the self-expressive aspect of purchases might help in some cases to modify or rewrite one's life narrative developed through socialization, getting free of stories and expectations one no longer wants to identify with (Shankar, Elliott, & Fitchett, 2009). Another way to look at the subject is to see consumption as a mean for sharing, giving and scarifying of what one obtained before. This approach has not only been practiced in religious institutions throughout time, but also researches as Belk (1988) explained that this could be, as Erich Fromm suggested, the key to realize one's identity and maximize self-expression without the threat of losing the object of our identity.

The majority of our society is still in the having and doing mode rather than sharing. This is why the results of this study also matter to policy makers. They can be an important inspiration for improving recommendations to consumers on how to best spend their disposable income to sustainably increase their happiness. Similarly, well-being in economically rich societies where a rise in income does no longer make a difference in happiness, could be increased. Policy makers could subsidy companies and brands that produce and sell more meaningful, environmentally friendly and sustainable products and experiences. Like that they could promote a shift in attention from fast consumption to a more thoughtful, slow consumption, supporting the definition of an environmentally friendly consumer identity. Just thinking about the constantly increasing obsessiveness with travelling and spending holidays in wasteful and inefficient hotels, demonstrates this necessity. Making problems like that visible to consumers might force businesses to embrace a new way of doing things and lead consumers to reconsider what they actually need, as well as where and how they want to spend their time and disposable income. Complementary to that, cost/benefit analysis for different options could be developed in order to support consumers making their daily choices on a higher level, as proposed by Csikszentmihalyi (2000).

6. Limitations and Future Directions

Although the present study is based on ground theorizing it may overlook important effects that future research might discover in the field of the experiential advantage and self-expression. When applying the findings of this study to daily life, the following limitations need to be considered.

First, the sample of this study is not an accurate representation of the general population. As it consisted of relatively young and highly educated participants, the results of this study are only applicable to this segment of the population. Differences might be found when investigating an older or less educated sample. In addition, it was restricted geographically to a German-speaking sample from Germany, Austria and Switzerland. Conducting the study in other countries, which are less individualistic as central European countries might yield different results. Another limitation is that the experiential groups were not completely equal in number of participants as exclusion rates differed. Although the sample size was sufficiently big, it is not impossible that this might still have a distorting effect on the outcome. It is recommended to perform a similar study with a larger sample size for higher external consistency. As this research was an online experiment, it cannot be ruled out that external influences might have distracted participants or whether other factors distorted their responses to the survey. Internal consistency in such an experiment is not as high as it would be in a laboratory experiment. Another limitation is that although people were assigned to groups differing in purchase self-expression, their selected purchases were probably not always 100% high self-expressive or low self-expressive. The effects might have been larger under perfect conditions, but still there were found significant effects. A last limitation is that hypothesis 2(b) could not be confirmed with both measures, time of recall and ease of recall. Significant evidence was found that when high in self-expression, experiential purchases were more easily recalled than material purchases employing the first measure. This effect was not found to be significant in the second one. Nevertheless, this finding could

be an interesting point to thoroughly investigate in future research. There seems to be a difference in how long something actually takes and how people perceive it. It would be interesting to investigate this specific process in the current context. In total, in the present study some important effects behind the experiential advantage were found as being significant, such as the heavier use of experiential purchases in socialization with the motivation to convey one's true self to someone else. Yet, this study did not explicitly investigate why this is actually the case. Why do people prefer to talk about their experiences? One potential possibility might be their strong narrative nature. Future studies could address this assumption more thoroughly.

7. Conclusion

Ancient philosopher's thinking about happiness was based on observations of the human being and drawing conclusions from it. Today, research can improve our understanding of everyday life by making observations and explicitly testing assumptions. These reliable and replicable methods can help us to shed new light on processes, enrich the scientific discussion and propose ways to improve the way we are living to date. This work has contributed to all of these aspects. It suggested to look at the experiential advantage from a different perspective and consider the effects of memory which emerge from human socialization. To add further, it invited to reconsider the great potential that both material and experiential purchases can have on our happiness in the case that they express who we truly are as person.

What the world needs today is a shift from a fast and excessive consumption which was characteristic for previous centuries, to a more thoughtful behavior, centered around getting to know oneself better, making sense of one's life and finding meaning through contributing to a higher mean. This is a circle in itself. As a person starts to center their attention on their true needs and the motivation behind their consumption is to achieve their goals, they approach their true self and to a meaningful life that is defined by true enjoyment in Aristotle's sense of "eudaimonia" (Ross, 1908). They focus on assuring the well-being of those around them and the planet.

"The things we see are the same things that are inside us.

There is no reality but that which we have within us."

Demian, Hermann Hesse

References

Ahuvia, A. C. (2005). Beyond the extended self: Loved objects and consumers' identity narratives. *Journal of consumer research*, *32*(1), 171–184.

Ang, S., Lim, E., & Leong, S. (2015). *In Pursuit of Happiness: Effects of Mental Subtraction and Alternative Comparison*. Dordrecht, Springer Netherlands.

Bastos, W. (2019). Consumers gain equivalent levels of happiness from sharing about an experience and an object. *European Journal of Marketing*.

Bastos, W., & Brucks, M. (2017). How and Why Conversational Value Leads to Happiness for Experiential and Material Purchases. *Journal of Consumer Research*, *44*(3), 598–612.

Baumeister, R. F., & Newman, L. S. (1994). *How Stories Make Sense of Personal Experiences: Motives that Shape Autobiographical Narratives*. Thousand Oaks, CA, Sage Publications.

Belk, R. W. (1985). Materialism: trait aspects of living in the material world. *Journal of Consumer research*, *12*(3), 265–280.

Belk, R. W. (1988). Possessions and the extended self. *Journal of consumer research*, *15*(2), 139–168.

Bhattacharjee, A., & Mogilner, C. (2013). Happiness from Ordinary and Extraordinary Experiences. *Journal of Consumer Research*, *41*(1), 1–17.

Biswas-Diener, R., & Dean, B. (2007). *Positive psychology coaching: Putting the science of happiness to work for your clients*. John Wiley & Sons.

Bruni, L. (2009). Frey, B.: Happiness. A revolution in economics. *Journal of Economics*, *98*(2), 177.

Caprariello, P. A., & Reis, H. T. (2013). To do, to have, or to share? Valuing experiences over material possessions depends on the involvement of others. *Journal of personality and social psychology*, *104*(2), 199.

Carroll, B. A., & Ahuvia, A. C. (2006). Some antecedents and outcomes of brand love. *Marketing letters*, *17*(2), 79–89.

Carter, T. J., & Gilovich, T. (2010). The relative relativity of material and experiential purchases. *Journal of personality and social psychology*, *98*(1), 146.

Carter, T. J., & Gilovich, T. (2012). I am what I do, not what I have: The differential centrality of experiential and material purchases to the self. *Journal of personality and social psychology*, *102*(6), 1304.

Chaplin, L. N., Lowrey, T. M., Ruvio, A. A., Shrum, L., & Vohs, K. D. (2020). Age differences in children's happiness from material goods and experiences: The role of memory and theory of mind. *International Journal of Research in Marketing*.

Clark, A. E. (2017). Happiness, income and poverty. *International Review of Economics*, *64*(2), 145–158.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience* (Vol. 1990). Harper & Row New York.

Csikszentmihalyi, M. (2000). The Costs and Benefits of Consuming. *Journal of Consumer Research*, *27*(2), 267–272.

Csikszentmihalyi, M., & Halton, E. (1981). *The meaning of things: Domestic symbols and the self*. Cambridge University Press.

Dai, H., Chan, C., & Mogilner, C. (2019). People Rely Less on Consumer Reviews for Experiential than Material Purchases. *Journal of Consumer Research*, *46*(6), 1052–1075.

Diener, E. (2009). Subjective well-being. In *The science of well-being* (pp. 11–58). Springer.

Diener, E., & Seligman, M. E. P. (2002). *Very Happy People*. Los Angeles, CA, SAGE Publications.

Dittmar, H., Bond, R., Hurst, M., & Kasser, T. (2014). The relationship between materialism and personal well-being: A meta-analysis. *Journal of personality and social psychology*, *107*(5), 879.

Duan, J., & Dholakia, R. R. (2018). *How Purchase Type Influences Consumption-Related Posting Behavior on Social Media: The Moderating Role of Materialism*. Routledge.

Dunn, E. W., Gilbert, D. T., & Wilson, T. D. (2011). If money doesn't make you happy, then you probably aren't spending it right. *Journal of Consumer Psychology*, *21*(2), 115–125.

Easterlin, R. A. (1974). Does economic growth improve the human lot? Some empirical evidence. In *Nations and households in economic growth* (pp. 89–125). Elsevier.

Epstein, S. (1994). *Integration of the Cognitive and the Psychodynamic Unconscious*. Arlington, Va., American Psychological Association.

Frankl, V. E. (1985). *Man's search for meaning*. Simon and Schuster.

Gallo, I., Sood, S., & Mann, T. C. (2017). *The Heart and the Head: On Choosing Experiences Intuitively and Possessions Deliberatively*.

Gallo, I., Townsend, C., & Alegre, I. (2019). Experiential product framing and its influence on the creation of consumer reviews. *Journal of Business Research*, *98*, 177–190.

Gardial, S. F., & Biehal, G. J. (1985). Memory accessibility and task involvement as factors in choice. *ACR North American Advances*.

Gilovich, T., & Gallo, I. (2020). Consumers' pursuit of material and experiential purchases: A review. *Consumer Psychology Review*, *3*(1), 20–33.

Gilovich, T., & Kumar, A. (2015). We'll always have Paris: The hedonic payoff from experiential and material investments. *Advances in experimental social psychology*, *51*, 147–187.

Gilovich, T., Kumar, A., & Jampol, L. (2015a). The beach, the bikini, and the best buy: Replies to Dunn and Weidman, and to Schmitt, Brakus, and Zarantonello. *Journal of Consumer Psychology*, *25*(1), 179–184.

Gilovich, T., Kumar, A., & Jampol, L. (2015b). A wonderful life: Experiential consumption and the pursuit of happiness. *Journal of Consumer Psychology*, *25*(1), 152–165.

Goldman, B. M., & Kernis, M. H. (2002). The role of authenticity in healthy psychological functioning and subjective well-being. *Annals of the American Psychotherapy Association*, *5*(6), 18–20.

Goodman, J. K., Malkoc, S. A., & Stephenson, B. L. (2016). Celebrate or commemorate? A material purchase advantage when honoring special life events. *Journal of the Association for Consumer Research*, *1*(4), 497-508.

Guevarra, D. A., & Howell, R. T. (2015). To have in order to do: Exploring the effects of consuming experiential products on well-being. *Journal of Consumer Psychology*, *25*(1), 28–41.

Habib, R., & Nyberg, L. (2007). Neural Correlates of Availability and Accessibility in Memory. *Cerebral Cortex*, *18*(7), 1720–1726.

Hayase, K., & Ura, M. (2015). *Ownership or Taking Action: Which Is More Important for Happiness?*

Hershfield, H. E., Mogilner, C., & Barnea, U. (2016). *People Who Choose Time Over Money Are Happier*. Los Angeles, CA, SAGE Publications.

Holt, D. B. (1997). Poststructuralist lifestyle analysis: Conceptualizing the social patterning of consumption in postmodernity. *Journal of Consumer research*, *23*(4), 326–350.

Hood, B. (2019). *Possessed: Why we want more than we need*. Oxford University Press.

Hornik, A., & Diesendruck, G. (2017). Extending the self via experiences: Undermining aspects of one's sense of self impacts the desire for unique experiences. *Social Cognition*, *35*(2), 181–203.

Howell, R. T., & Hill, G. (2009). The mediators of experiential purchases: Determining the impact of psychological needs satisfaction and social comparison. *The Journal of Positive Psychology*, *4*(6), 511–522.

IATA. (2020). *Anzahl der Flüge in der weltweiten Luftfahrt von 2014 bis 2019 (in Millionen) [Graph]*. https://de-statista-com.eaccess.ub.tum.de/statistik/daten/studie/411620/umfrage/anzahl-der-weltweiten-fluege/.

Kasser, T. (2000). *The High Price of Materialism*. MIT Press.

Kasser, T., & Ryan, R. M. (1996). Further examining the American dream: Differential correlates of intrinsic and extrinsic goals. *Personality and social psychology bulletin*, *22*(3), 280–287.

Kernis, M. H., & Goldman, B. M. (2006). A multicomponent conceptualization of authenticity: Theory and research. *Advances in experimental social psychology*, *38*, 283–357.

Keven, N. (2016). *Events, narratives and memory*. Dordrecht, Springer Netherlands.

Kifer, Y., Heller, D., & Perunovic, W. Q. E. (2013). *The Good Life of the Powerful: The Experience of Power and Authenticity Enhances Subjective Well-Being*. Los Angeles, CA, SAGE Publications.

Kim, H. S., & Ko, D. (2007). Culture and self-expression. *The self*, 325–342.

Kim, J., Seto, E., Christy, A. G., & Hicks, J. A. (2016). Investing in the real me: Preference for experiential to material purchases driven by the motivation to search for true self-knowledge. *Self and Identity*, *15*(6), 727–747.

Kircher, T., & David, A. (2003). *The self in neuroscience and psychiatry*. Cambridge University Press.

Kulkofsky, S., Wang, Q., & Ceci, S. J. (2008). Do better stories make better

memories? Narrative quality and memory accuracy in preschool children. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *22*(1), 21–38.

Kumar, A., & Gilovich, T. (2015). Some "thing" to talk about? Differential story utility from experiential and material purchases. *Personality and Social Psychology Bulletin*, *41*(10), 1320–1331.

Law, J. R., Flanery, M. A., Wirth, S., Yanike, M., Smith, A. C., Frank, L. M., ... Stark, C. E. (2005). Functional magnetic resonance imaging activity during the gradual acquisition and expression of paired-associate memory. *Journal of Neuroscience*, *25*(24), 5720–5729.

Lee, J. C., Hall, D. L., & Wood, W. (2018). Experiential or Material Purchases? Social Class Determines Purchase Happiness. *Psychological Science*, *29*(7), 1031–1039.

Lenoir, F. (2015). *Happiness: A Philosopher's Guide*. Melville House Pub.

Lenton, A. P., Bruder, M., & Slabu, L. (2013). *How Does "Being Real" Feel? The Experience of State Authenticity*.

Lindgren, A., Vanhamme, J., & Beverland, M. B. (2009). *Memorable customer experiences: A research anthology*. Gower Publishing.

Lyubomirsky, S., King, L., & Diener, E. (2005). The benefits of frequent positive affect: Does happiness lead to success? *Psychological bulletin*, *131*(6), 803.

Marsh, E. J., & Tversky, B. (2004). Spinning the stories of our lives. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *18*(5), 491–503.

Matz, S. C., Gladstone, J. J., & Stillwell, D. (2016). Money buys happiness when spending fits our personality. *Psychological science*, *27*(5), 715–725.

Mitchell, T. R., Thompson, L., Peterson, E., & Cronk, R. (1997). Temporal adjustments in the evaluation of events: The "rosy view". *Journal of experimental social psychology*, *33*(4), 421–448.

Nicolao, L., Irwin, J. R., & Goodman, J. K. (2009). Happiness for Sale: Do Experiential Purchases Make Consumers Happier than Material Purchases? *Journal of Consumer Research*, *36*(2), 188–198.

Pavot, W., & Diener, E. (2008). The satisfaction with life scale and the emerging construct of life satisfaction. *The journal of positive psychology*, *3*(2), 137–152.

Payne, J., Bettman, J., & Johnson, E. (1991). Consumer decision making. *Handbook of consumer behaviour*, 50–84.

Pchelin, P., & Howell, R. T. (2014). The hidden cost of value-seeking: People do not accurately forecast the economic benefits of experiential purchases. *The Journal of Positive Psychology*, *9*(4), 322–334.

Peng, Z., & Ye, M. (2015). An introduction of purchase types and happiness. *Journal of Service Science and Management*, *8*(01), 132.

Reed II, A., Forehand, M. R., Puntoni, S., & Warlop, L. (2012). Identity-based consumer behavior. *International Journal of Research in Marketing*, *29*(4), 310–321.

Richins, M. L. (2004). The material values scale: Measurement properties and development of a short form. *Journal of consumer Research*, *31*(1), 209–219.

Richins, M. L. (2012). When Wanting Is Better than Having: Materialism, Transformation Expectations, and Product-Evoked Emotions in the Purchase Process. *Journal of Consumer Research*, *40*(1), 1–18.

Rosenzweig, E., & Gilovich, T. (2012). *Buyer's Remorse or Missed Opportunity? Differential Regrets for Material and Experiential Purchases*. American Psychological Association.

Ross, W. (1908). *Aristotle: nicomachean ethics. Translated by WD Ross* (Vol. 15). Oxford, Clarendon Press.

Ryan, R. M., LaGuardia, J. G., & Rawsthorne, L. J. (2005). Self-complexity and the authenticity of self-aspects: Effects on well being and resilience to stressful events. *North American Journal of Psychology*, *7*(3).

Sartre, J.-P. (2015). Being and nothingness. *Central Works of Philosophy v4: Twentieth Century: Moore to Popper*, *4*, 155.

Schultz, S. E., Kleine, R. E., & Kernan, J. B. (1989). 'These are a few of my favorite things': Toward an explication of attachment as a consumer behavior construct. *Advances in consumer research*, *16*(1), 359–366.

Shankar, A., Elliott, R., & Fitchett, J. A. (2009). Identity, consumption and narratives of socialization. *Marketing Theory*, *9*(1), 75–94.

Sheldon, K. M., Ryan, R. M., Rawsthorne, L. J., & Ilardi, B. (1997). Trait self and true self: Cross-role variation in the Big-Five personality traits and its relations with psychological authenticity and subjective well-being. *Journal of personality and social psychology*, *73*(6), 1380.

Shrum, L. J., Lowrey, T. M., & Pandelaere, M. (2014). *Materialism: the good, the bad, and the ugly*. Routledge.

Shrum, L. J., Wong, N., Arif, F., Chugani, S. K., Gunz, A., Lowrey, T. M., ... Ruvio, A. (2013). Reconceptualizing materialism as identity goal pursuits: Functions, processes, and consequences. *Journal of Business Research*, *66*(8), 1179–1185.

Toffler, A. (1970). *Future shock*. Bantam.

Tully, S. M., Hershfield, H. E., & Meyvis, T. O. M. (2015). Seeking Lasting Enjoyment with Limited Money: Financial Constraints Increase Preference for Material Goods over Experiences. *Journal of Consumer Research*, *42*(1), 59–75.

Tulving, E. (1972). Episodic and semantic memory. *Organization of memory*, *1*, 381–403.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, *5*(2), 207–232.

Van Boven, L., & Gilovich, T. (2003). To do or to have? That is the question. *Journal of personality and social psychology*, *85*(6), 1193.

Vignoles, V. L., Regalia, C., & Manzi, C. (2006). *Beyond Self-Esteem: Influence of Multiple Motives on Identity Construction*. American Psychological Association.

Wallendorf, M., & Arnould, E. J. (1988). "My favorite things": A cross-cultural inquiry into object attachment, possessiveness, and social linkage. *Journal of Consumer Research*, *14*(4), 531–547.

Weidman, A. C., & Dunn, E. W. (2016). The unsung benefits of material things: Material purchases provide more frequent momentary happiness than experiential purchases. *Social Psychological and Personality Science*, *7*(4), 390–399.

Weingarten, E., & Goodman, J. K. (2020). Re-examining the Experiential Advantage in Consumption: A Meta-Analysis and Review. *Journal of Consumer Research*.

Weinswig, D. (2016). *Millennials Go Minimal: The Decluttering Lifestyle Trend That Is Taking Over*. https://www.forbes.com/sites/deborahweinswig/2016/09/07/millennials-go-minimal-the-decluttering-lifestyle-trend-that-is-taking-over/?sh=57a86cfd3755. (Accessed: 2020-12-29)

Westbrook, R. A. (1987). Product/consumption-based affective responses and postpurchase processes. *Journal of marketing research*, *24*(3), 258–270.

Wirth, S., Yanike, M., Frank, L. M., Smith, A. C., Brown, E. N., & Suzuki, W. A. (2003). Single neurons in the monkey hippocampus and learning of new associations. *Science*, *300*(5625), 1578–1581.

Wood, A. M., Linley, P. A., Maltby, J., Baliousis, M., & Joseph, S. (2008). The authentic personality: A theoretical and empirical conceptualization and the development of the Authenticity Scale. *Journal of Counseling Psychology*, *55*(3), 385.

Wyer, R. S., & Srull, T. K. (1986). *Human Cognition in Its Social Context*. American Psychological Association.

Yu, Y., Jing, F., Su, C. T., Zhou, N., & Nguyen, B. (2016). Impact of material vs. experiential purchase types on happiness: The moderating role of self-discrepancy. *Journal of Consumer Behaviour*, *15*(6), 571–579.

Zhang, J. W., Howell, R. T., Caprariello, P. A., & Guevarra, D. A. (2014). Damned if they do, damned if they don't: Material buyers are not happier from material or experiential consumption. *Journal of Research in Personality*, *50*, 71–83.

Junior Management Science 8(1) (2023) 188-218

Junior Management Science

journal homepage: www.jums.academy

Critical analysis of the VAT treatment of vouchers since the annual Tax Act 2018

Kritische Analyse der umsatzsteuerlichen Behandlung von Gutscheinen seit dem Jahressteuergesetz 2018

Carina Menhofer

Hochschule für angewandte Wissenschaften Augsburg

Abstract

Vouchers are used in many different ways in numerous sectors and, in addition to the classic target group of end consumers, can now be observed throughout the entire service chain. The issuance of vouchers enables companies to generate liquid funds without providing any service immediately, so they have become very important for the economy. In the past, the VAT treatment of vouchers was characterized by legal uncertainties due to a lack of harmonization. The so-called Voucher Directive was intended to standardize the VAT treatment of vouchers within the EU member states. The aim of this paper is to analyze the implementation of the Voucher Directive into national law with the Annual Tax Act 2018 as well as the corresponding letter of the German Ministry of Finance with regard to its application in practice. In doing so, the scientific literature, case law, legislation and administrative regulations on the VAT treatment of vouchers are consulted. The work gives an overview of the tax treatment and points out problem areas as well as structuring possibilities. In this context, the temporary VAT reduction in connection with the Corona crisis and its effects are also addressed. The results show that an European regulation on the VAT treatment of vouchers is in principle to be viewed positively. However, numerous legal uncertainties remain with regard to the classification of vouchers and its assessment in general.

Zusammenfassung

Gutscheine werden in zahlreichen Branchen mannigfach verwendet und sind neben der klassischen Zielgruppe Endverbraucher inzwischen innerhalb der gesamten Leistungskette zu beobachten. Die Ausgabe von Gutscheinen ermöglicht den Unternehmen liquide Mittel zu generieren, ohne eine sofortige Gegenleistung erbringen zu müssen, sodass ihnen in der Wirtschaft eine große Bedeutung zugekommen ist. Die umsatzsteuerliche Behandlung von Gutscheinen war in der Vergangenheit, aufgrund fehlender Harmonisierung geprägt von Rechtsunsicherheiten. Mit der sogenannten Gutschein-Richtlinie sollte die umsatzsteuerliche Behandlung von Gutscheinen innerhalb der EU-Mitgliedstaaten vereinheitlicht werden. Ziel der vorliegenden Arbeit ist es, die Umsetzung der Gutschein-Richtlinie in nationales Recht mit Jahressteuergesetz 2018 sowie dem entsprechenden BMF-Schreiben auf seine Anwendung in der Praxis zu analysieren. Dabei wird die wissenschaftliche Literatur, Rechtsprechung, Gesetzgebung und Verwaltungsvorschriften zur umsatzsteuerlichen Behandlung von Gutscheinen herangezogen. Die Arbeit gibt einen Überblick über die steuerliche Behandlung und zeigt sowohl Problemfelder als auch Gestaltungsmöglichkeiten auf. Hierbei wird auch auf die temporäre Umsatzsteuersenkung im Zusammenhang mit der Corona Krise und deren Auswirkungen eingegangen. Die Ergebnisse zeigen, dass eine unionsübergreifende Regelung zur umsatzsteuerlichen Würdigung von Gutscheinen grundsätzlich positiv zu werten ist. Jedoch bleiben in Bezug auf die Klassifizierung der Gutscheine als auch der Würdigung weiterhin zahlreiche Rechtsunsicherheiten bestehen.

Keywords: Umsatzsteuer; Einzweckgutschein; Mehrzweckgutschein; Gutschein-Richtlinie.

1. Problemstellung

In der vorliegenden Arbeit wird die umsatzsteuerliche Behandlung von Gutscheinen seit dem Jahressteuergesetz 2018 kritisch analysiert.

Gutscheine werden in zahlreichen Branchen mannigfach verwendet. Charakteristisch ist der Einzelhandel, wo der

DOI: https://doi.org/10.5282/jums/v8i1pp188-218

Endverbraucher mit diversen Ausprägungen dieses Marketinginstruments – wie beispielsweise Prämien-, Rabatt-, Wert- oder Sachgutscheine – konfrontiert wird. Die Beweggründe der Unternehmen zur Implementation dieses Marketinginstruments sind Neukundengewinnung bzw. Stammkundenpflege sowie Absatzförderung.[1] Gutscheine etablieren sich seit Jahren als ideales Geschenk für diverse Anlässe, wobei Unternehmen beim Absatz von Wert- und Sachgutscheine umgehend von einer sofortigen Vereinnahmung des Entgelts profitieren.[2]

Weit größere Popularität wurde den Gutscheinen im Laufe der Corona-Krise zuteil. Einerseits wurden sie als Rettungsinstrument zur Aufrechterhaltung der Liquidität lokaler Gastronomiebetriebe und Einzelhändler ausgerufen[3], zum anderen wurde der Gutschein - statt einer baren Rückgewähr - für Anzahlungen abgesagter Veranstaltungen und Pauschalreisen vom Bundeskabinett beschlossen.[4] Neben der klassischen Zielgruppe Endverbraucher bzw. Privatperson ist der Verkauf und Einsatz von Gutscheinen als Anreiz- und Vertriebsinstrument inzwischen innerhalb der gesamten Leistungskette zu beobachten.[5] Des Weiteren erfreut sich der Gutschein als Möglichkeit zur Entgeltoptimierung bei Mitarbeitern großer Beliebtheit.[6]

Bereits im Jahr 2012 bezifferte die EU-Kommission in einer Pressemitteilung das ökonomische Gewicht von Gutscheinen mit mehr als 52 Mrd. EUR.[7] Trotz der enormen wirtschaftlichen Kraft gab es bis zum 31.12.2018 weder eine nationale noch unionsrechtlich einheitliche gesetzliche Vorschrift zu deren umsatzsteuerlichen Würdigung. Folglich etablierten die EU-Mitgliedstaaten nationale Lösungsansätze, welche unvermeidlich zu Missverhältnissen der Besteuerung, wie der Doppelbesteuerung oder auch Nichtbesteuerung, beigetragen haben.[8] Erschwerend wurden zudem seitens des EuGHs[9] konträre Feststellungen zur umsatzsteuerlichen Behandlung von Gutscheinen getroffen[10], welche durch die deutsche Finanzverwaltung für die nationale Rechtspraxis teilweise auch abgelehnt wurde.[11]

Eine unionsweite einheitliche Regelung zur Besteuerung von Gutscheinen wurde mit der sog. Gutschein-Richtlinie gemäß Art. 30a, 30b MwStSystRL sowie Art. 73a MwStSystRL festgesetzt und zum 1.1.2019 in nationales Recht umgesetzt.[12] Die bisherige umsatzsteuerliche Würdigung von Gutscheinen kann noch nicht vollständig ad acta gelegt werden, da die gesetzliche Neuregelung gemäß § 27 Abs. 23 UStG erst auf nach dem 31.12.2018 ausgestellte Gutscheine Anwendung findet.

Mannigfaltige Beiträge in der Literatur als auch Eingaben der Verbände warfen schon während des Gesetzgebungsverfahrens Einzelfragen hinsichtlich der nationalen Umsetzung der Gutschein-Richtlinie in der Praxis auf.[13] Diesbezügliche Hilfestellungen wurden seitens der Finanzverwaltung am 18.12.2019 als Entwurfsschreiben an einige Verbände zur Stellungnahme zugesandt.[14] Am 2.11.2020 wurde ein BMF-Schreiben[15] zur umsatzsteuerrechtlichen Behandlung von Einzweck- und Mehrzweck-Gutscheinen veröffentlicht, welches mit zahlreichen Anwendungsbeispielen versehen ist. Jedoch erschließen sich hiermit nicht alle mit Gutscheinen in Zusammenhang stehenden Fragestellungen und ist auf ei-

[1] Vgl. Oliver L. Knöfel, „Gift Cards (Geschenkkarten) im Einzelhandel: Rechtsnatur, Funktionsweise, Probleme." WM, Nr. 18 (2017): 833–840, S. 834.

[2] Vgl. Gerhard Michel, „Internet-Marktplätze und Gutscheine im Fokus der Steuergesetzgebung - umsatzsteuerrechtliche Änderungen zum 01.01.2019." jM (2018): 431–437, S. 434.

[3] Vgl. Institut für Wissen in der Wirtschaft GmbH, „Umsatzsteuer: Corona-Erleichterungen für Gastronomie und die Auswirkung auf Gutscheinverkäufe." AStW (2020): 473–474, S. 473.

[4] Vgl. „Gesetz zur Abmilderung der Folgen der COVID-19-Pandemie im Veranstaltungsrecht und im Recht der Europäischen Gesellschaft (SE) und der Europäischen Genossenschaft (SCE)." BGBl. I 2020, Nr. 22: 948– 950; „Gesetz zur Abmilderung der Folgen der COVID-19-Pandemie im Pauschalreisevertragsrecht und zur Sicherstellung der Funktionsfähigkeit der Kammern im Bereich der Bundesrechtsanwaltsordnung, der Bundesnotarordnung, der Wirtschaftsprüferordnung und des Steuerberatungsgesetztes während der COVID-19-Pandemie." BGBl. I, Nr. 35 (2020): 1643-1647.

[5] Vgl. Kathrin Feil, Nina Kupke und Eva Greisl, „Aktuelle Entwicklungen bei der umsatzsteuerlichen Behandlung von Gutscheinen." BB (2012): 3113–3123, S. 3113.

[6] Vgl. Hans Herrler und Caroline Herrler, „Mehr Netto vom Brutto." NWB, Nr. 6 (2019): 350–364, S. 357.

[7] Vgl. Europäische Kommission, „Kommission schlägt neue MwSt-Vorschriften für Gutscheine vor.". https://ec.europa.eu/commission/presscorner/detail/de/IP_12_464 (letzter Zugriff: 21. Februar 2021).

[8] Vgl. Europäische Kommission, „COM (2012) 206 final: Vorschlag für eine Richtlinie des Rates zur Änderung der Richtlinie 2006/112/EG über das gemeinsame Mehrwertsteuersystem hinsichtlich der Behandlung von Gutscheinen.". https://ec.europa.eu/transparency/regdoc/rep/1/2012/DE/1-2012-206-DE-F1-1.Pdf (letzter Zugriff: 4. Februar 2021), S. 2.

[9] Vgl. EuGH, „Urt. v. 29.7.2010 - Rs. C 40/09 - Astra Zeneca." DStR (2010): 1623-1625; EuGH, „Urt. v. 16.12.2010 - Rs. C-270/09 - MacDonald Resorts Ltd." UR (2011): 462–470.

[10] Vgl. Suse Martin, „§ 3 UStG." In Sölch/Ringleb Umsatzsteuergesetz (München: C.H. Beck oHG), 89. EL Juni 2020:Rz. 54.

[11] Vgl. BMF, „Schr. v. 20.12.2010 - IV D 2-S 7200/07/10011/002." DStR (2010): 629.

[12] Vgl. Richtlinie (EU) 2016/ 1065 v. 27.6.2016.

[13] Vgl. Stefan Heinrichshofen und Nina Kupke, „Gutscheine nach der Änderung der Mehrwertsteuer- Systemrichtlinie (MwStSystRL) - Anmerkungen zu den Auswirkungen auf die deutsche Besteuerungspraxis: (Teil 1)." UVR (2017): 142–148, S. 147; Philipp J. Thiele und Isabelle König, „Die umsatzsteuerrechtliche Behandlung von Gutscheinen: Eine kritische Betrachtung der Neuregelungen." UR (2018): 933–940, S. 938; Kathrin Feil und Vivien Polok, „Hält die EU-Gutschein-Richtlinie was sie verspricht?" MwStR (2017): 65-68, S. 68; Deutscher Industrie- und Handeslkammertag e.V. et al., „Stellungnahme zum Umsetzungsvorschlag der Gutschein-Richtline (EU) 2016/1065 durch die Finanzverwaltung: Gz: III C 2 - S7100/07/10045:003; DOK 2017/0597914.". https://www.gdv.de/resource/blob/5720/c147f626743d08d287241a9a30a4fc55/zum-umsetzungsvorschlag-der-gutschein-richtlinie-1309327852-data.pdf (letzter Zugriff: 28. Dezember 2020); DEUTSCHER STEUERBERATERVERBAND e.V., „Referentenentwurf eines Jahressteuergesetzes 2018.". https://www.dstv.de/interessenvertretung/steuern/stellungnahmen-steuern/2018-s07-refentwurf-jstg2018 (letzter Zugriff: 19. Mai 2020); Bundessteuerberaterkammer, „Diskussionsentwurf für eine gesetzliche Umsetzung der Gutschein-Richtlinie (EU) 2016/1065.". https://www.bstbk.de/downloads/bstbk/presse-und-kommunikation/stellungnahmen/BStBK_2017-019_Diskussionsentwurf-gesetzliche-Umsetzung-Gutschein_1.pdf(letzterZugriff:1.Februar2021).

[14] Vgl. Axel Leonard und Martin Robisch, „§ 3 UStG." In Bunjes UStG (München: C.H. Beck oHG), Rn. 321.

[15] BMF, „Schr. v. 2.11.2020 - III C 2 - S 7100/19/10001 :002, DOK 2020/1051603." BStBl. I (2020): 1121.

ne praxistaugliche Umsetzung zu analysieren. Beinahe zwei Jahre umfasste der Zeitraum zwischen Einführung des Gesetzes, Entwurf eines BMF- Schreibens, der anschließenden Stellungnahmen und des finalen BMF-Schreibens, was durchaus die Komplexität der Thematik widerspiegelt. Daneben wirken offene Fragen in Zusammenhang mit der temporären Umsatzsteuersatzsenkung vom 1.7.2020 bis 31.12.2020 und der Ausgabe von Gutscheinen nach.

2. Wirtschaftliche Betrachtungsweise von Gutscheinen

2.1. Hintergrund

Auf europäischer Ebene geht man von einem jährlichen Gutscheinvolumina von 52 Mrd. EUR aus, wobei vorbezahlte Telekommunikationsdienste fast 70 % des Gutscheinmarktes beanspruchen, gefolgt von Geschenkgutscheinen und Rabattgutscheinen.[16] Gemäß der Christmas Surveys von Deloitte profitierte insbesondere das Weihnachtsgeschäft in den letzten Jahren von stetiger Beliebtheit der Gutscheine. Lediglich im Jahr 2020 lässt sich ein gegenläufiger Trend erkennen.20 Eine Untersuchung der GfK hinsichtlich des Weihnachtsgeschäfts in Deutschland 2017 erfasste durch den Kauf von Gutscheinen ein Volumen in Höhe von 1,84 Mrd. EUR.[17]

Gutscheine sind bei Schenkern als auch bei Beschenkten nicht nur zu Weihnachten, sondern zu jeglichen Anlässen, populär. Dabei hat das Spektrum der Geschenkgutscheine in den letzten Jahren vielfach zugenommen, wodurch sich neue Vertriebs- und Wirtschaftszweige etablieren konnten. Zu beobachten ist zum einen ein gestiegener gewerblicher Handel mit Gutscheinen fremder Anbieter (z.B. Gutscheinkarte von Apple im Kaufhaus oder an der Tankstelle)[18] und zum anderen konnte sich in den letzten Jahren in der Vermittlung von Erlebnisgutscheinen ein komplett neuer Geschäftszweig entwickeln.[19] Namentliche Unternehmen wie „mydays" oder „Jochen Schweitzer" profitieren von einem in 2017 geschätzten Marktvolumen für Erlebnisgeschenke von bis zu 500 Mio. EUR in Deutschland.[20]

2.2. Erscheinungsformen

Gutscheine werden mit unterschiedlichen wirtschaftlichen Zielsetzungen generiert, wobei der Begriff „Gutschein" im Alltag als Sammelbegriff zu verstehen ist.[21] Der Duden definiert einen Gutschein als einen „Schein, der den Anspruch auf eine bestimmte Sache, auf Waren oder Dienstleistungen mit einem bestimmten Gegenwert bestätigt".[22] Durch den Einsatz des Instruments Gutschein soll – egal in welcher Erscheinungsform – dem Unternehmer eine Absatzförderung generiert werden. Dabei kann die Ausgestaltung von „Gutscheinen" unterschiedlich erfolgen, wobei aus rechtlicher und betriebswirtschaftlicher Sicht der Verpflichtungsgrad des ausgebenden Unternehmers eine zentrale Rolle einnimmt.[23] Aufgrund der stetigen Veränderung hinsichtlich der wirtschaftlichen Interessen sind die folgenden Erscheinungsformen als nicht abschließend anzusehen.

2.2.1. Eigenleistungsgutscheine

Werbegutscheine werden genutzt, um potenzielle Kunden zum Kaufabschluss zu bewegen, daher sind sie vergleichbar mit Werbeflyer, welche zum Besuch eines Geschäftes oder Onlineshops animieren.[24] Charakteristisch für Info-Gutscheine, die beispielsweise einen kostenlosen Bezug von Informationen über ein Unternehmen und dessen Produkte ermöglichen, ist ein fehlendes Leistungsversprechen. Oftmals werden Werbegutscheine auch als Produktgutschein generiert, welcher einen kostenlosen Bezug bestimmter Waren oder Leistungen ermöglicht. Durch den Abdruck eines solchen Gutscheins in einer Zeitschrift oder Broschüre richtet der Unternehmer ein verbindliches Angebot an einen unbestimmten, aber bestimmbaren Personenkreis. In Folge darf der werbende Unternehmer, soweit keine qualitativen Schranken („gilt bis zum…"; „die ersten…) gesetzt wurden, die Leistung nicht verweigern.[25]

Nachlassgutscheine bzw. Coupons sind Marketinginstrumente zur Schaffung einer zielgerichteten Verkaufsförderung durch einen prozentualen oder betragsmäßigen Preisnachlass.[26] Zum einen können Coupons über die unterschiedlichsten Kanäle, wie z.B. Postwurfsendungen, Couponblätter in Zeitschriften oder über Apps, an den Endverbraucher herangetragen werden.[27] Zum anderen sind beim Unternehmen gesammelte Treuepunkte aus vorangegangen Einkäufen ein verbreitetes Instrument, um Preisnachlässe zu gewähren. Mit der Vorlage des Nachlassgutscheins unterbreitet der Kunde dem Unternehmen ein Angebot, Waren oder Leistungen zu einem geringeren Preis abzunehmen, welcher vom Unternehmer angenommen werden kann. Theoretisch kann der Unternehmer diesen Nachlassgutschein jedoch auch ablehnen.

[16] Vgl. Europäische Kommission, „Kommission schlägt neue MwSt-Vorschriften für Gutscheine vor".

[17] Vgl. GfK, „Was lag unter dem Weihnachtsbaum? Ein Rückblick auf das Weihnachtsgeschäft 2017.". https://www.gfk.com/de/insights/was-lag-unter-dem-weihnachtsbaum (letzter Zugriff: 4. Februar 2021).

[18] Vgl. Feil, Kupke und Greisl, „Aktuelle Entwicklungen bei der umsatzsteuerlichen Behandlung von Gutscheinen", S. 3113.

[19] Vgl. Boost Internet GmbH, „Erlebnisgeschenke-Report: Eine Studie zum Markt für Erlebnisgeschenke in Deutschland.". http://www.erlebnisgeschenke.de/download/marktstudie.pdf (letzter Zugriff: 3. März 2021), S. 9; Bundessteuerberaterkammer, „BMF-Schreiben zur Anwendung der Neuregelungen zu den Gutscheinen.". https://www.bstbk.de/downloads/bstbk/presse-und-kommunikation/stellungnahmen/BStBK_2019-006_Stellungn_USt_Gutscheine.pdf (letzter Zugriff: 3. Februar 2021), S. 2.

[20] Vgl. Boost Internet GmbH, „Erlebnisgeschenke-Report", S. 8.

[21] Vgl. Martin Ahrens, „Gutscheine." BB (1996): 2477–2481, S. 2477.

[22] Duden - Deutsches Universalwörterbuch, 9., vollständig überarbeitete und erweiterte Auflage (Berlin: Dudenverlag, 2019).

[23] Vgl. Claus Koss, „Gutscheine und ihre rechtliche Beurteilung." DB, Nr. 47 (2019): 2593-2599, S. 2594.

[24] Vgl. ebd., S. 2593-2594.

[25] Vgl. Ahrens, „Gutscheine", S. 2479.

[26] Vgl. Christoph Burmann et al., Kompakt-Lexikon Marketingpraxis: 2.200 Begriffe nachschlagen, verstehen, anwenden. SpringerLink (Wiesbaden: Springer Gabler, 2013), S. 116.

[27] Vgl. Alexandra Ranzinger, Praxiswissen Kundenbindungsprogramme: Konzeption und operative Umsetzung, 2., überarbeitete Auflage (Wiesbaden: Springer Gabler, 2017), S.4-5.

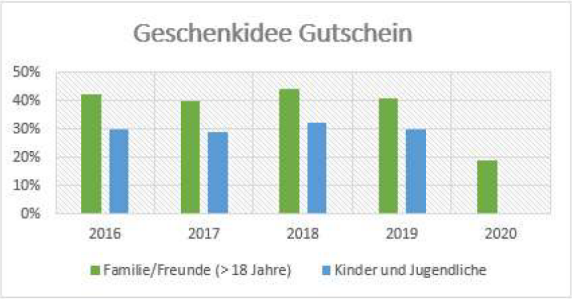

Abbildung 1: Gutscheine als beliebtes Geschenk zu Weihnachten.
Vgl. Deloitte GmbH Wirtschaftsprüfungsgesellschaft, „Christmas Survey 2020: COVID-19 halbiert das deutsche Weihnachtsbudget.".https://www2.deloitte.com/de/de/pages/consumer-business/articles/christmas-survey.html (letzter Zugriff: 24. Februar 2021); Deloitte GmbH Wirtschaftsprüfungsgesellschaft, „Christmas Survey 2019: Zentrale Ergebnisse für Deutschland.".https://www2.deloitte.com/de/de/pages/consumer-business/articles/christmas-survey.html (letzter Zugriff: 24. Februar 2021); Deloitte GmbH Wirtschaftsprüfungsgesellschaft, „Christmas Survey 2018: Zentrale Ergebnisse für Deutschland.".https://www2.deloitte.com/de/de/pages/consumer-business/articles/christmas-survey.html (letzter Zugriff: 24. Februar 2021); Deloitte GmbH Wirtschaftsprüfungsgesellschaft, „Christmas Survey 2017: Zentrale Ergebnisse für Deutschland.".https://www2.deloitte.com/de/de/pages/consumer-business/articles/christmas-survey.html (letzter Zugriff: 24. Februar 2021).

Folgen könnten sich jedoch aus wettbewerbsrechtlicher Sicht aufgrund von irreführender Werbung ergeben.[28]

Bei der genossenschaftlichen Rückvergütung handelt es sich um eine spezielle steuerliche Regelung, bei der ein Unternehmen in der Rechtsform einer eG Teile ihres Jahresergebnisses an ihre Mitglieder zurückvergütet. Dabei ist die genossenschaftliche Rückvergütung keine Gewinnverteilung, sondern wird aufgrund einer Leistungsbeziehung mit den Mitgliedern gewährt.[29] Historisch betrachtet stellen sie demnach eine Sonderform der Nachlassgutscheine dar und sind heute noch bei Einkaufs- und Leistungsgenossenschaften vorzufinden.[30]

Ein Leistungsgutschein kann in Form eines Nennwertgutscheins oder eines Warengutscheins auftreten. Sie ermöglichen Inhaber auf eine unbestimmte Leistung bis zu einem bestimmten Wert oder eine bestimmte Leistung zu erwerben. Rechtlich ist der Leistungsgutschein nach herrschender Meinung als sog. kleines Inhaberpapier gemäß § 807 BGB einzuordnen, da sich der Aussteller zu einer Leistung an den jeweiligen Inhaber verpflichtet. Aufgrund positiven Vorliegens eines zivilrechtlichen Anspruchs, unterliegt dieser der allgemeinen gesetzlichen Verjährungsvorschrift nach § 195 BGB.[31]

2.2.2. Fremdleistungsgutscheine

In den oben genannten Ausführungen ist davon auszugehen, dass der Aussteller selbst eine Leistung erbringt. Die Verpflichtung zur Erbringung der Leistung kann sich daraus ergeben, dass er mit dem Kunden einen Vertrag geschlossen hat oder diese aufgrund einer öffentlich-rechtlichen Verpflichtung zu erbringen ist. Abgrenzend zu den obenstehenden Ausführungen hat diese Leistung nicht vom Vertragspartner oder dem öffentlich-rechtlich dazu Verpflichteten, sondern nun von einem Dritten erbracht zu werden. Beispielhaft sind hier Warengutscheine für Asylbewerber zu nennen.[32] Der Fremdleistungsgutschein ist daher regelmäßig als Anweisung i.S.v. § 783 BGB einzuordnen.[33] Dadurch wird ein Dritter (Anweisungsempfänger) ermächtigt eine Leistung bei einer anderen Person (der Angewiesene) gegen Vorlage des Gutscheins in eigenem Namen zu erheben und den Angewiesenen ermächtigt, an den Empfänger für Rechnung des Ausstellers (der Anweisende) zu leisten.[34]

2.3. Gutscheine als Liquiditätsquelle

Neben dem Marketingaspekt ist der entscheidende wirtschaftliche Faktor bei der Ausgabe von Gutscheinen gegen Entgelt die sofortige Einnahme von liquiden Mitteln, wobei keine sofortige Leistung seitens des Unternehmers zu erbringen ist. Dem Grunde nach gewährt der Kunde dem Aussteller des Gutscheins einen zinslosen Kredit. Insbesondere vor dem Hintergrund, dass bis zu 30 % der Gutscheine von ihren Besitzern nicht eingelöst werden, ist der Verkauf von Gutscheinen ein gewinnbringendes Instrument.[35] Zudem erwirbt der Beschenkte Schätzungen zu folge bei Einlösung eines Gutscheins Waren zu einem höheren als auf dem Gutschein aufgedruckten Betrag.[36]

[28]Vgl. Koss, „Gutscheine und ihre rechtliche Beurteilung", S. 2593-2594.
[29]Vgl. Thomas Strieder und Dieter Gahlen, „§ 336 HGB." In beck-online.GROSSKOMMENTAR, hrsg. von Martin Henssler (München: C.H. Beck oHG, 2020), Rn. 16.
[30]Vgl. Koss, „Gutscheine und ihre rechtliche Beurteilung", S. 2593.
[31]Vgl. ebd., S. 2595.
[32]Vgl. OLG Düsseldorf, „Urt. v. 09.08.1994 - 24 U 181/93." NVwZ (1995): 827-829, S. 827.
[33]Vgl. Ahrens, „Gutscheine", S. 2481.
[34]Vgl. Peter W. Heermann, „§ 783 BGB." In juris Praxiskommentar BGB: Band 2 - Schuldrecht, hrsg. von Markus Junker et al. 9. Auflage, Juris Praxiskommentar BGB / Gesamtherausgeber: Prof. Dr. em. Maximilian Herberger, Prof. Dr. Dr. Dr. h.c. mult. Michael Martinek M.C.J. (New York), Prof. em. Dr. Dr. h.c. mult. Helmut Rüßmann, Prof. Dr. Stephan Weth, Prof. Dr. Markus Würdinger Band 2 (Saarbrücken: Juris, 2020), Rn. 1.
[35]Vgl. Eilnachricht, „Liquidität: Mit Gutscheinen Zahlungsfähigkeit verbessern." nwb BB, Nr. 9 (2018): 260.
[36]Vgl. Knöfel, „Gift Cards (Geschenkkarten) im Einzelhandel", S. 834.

Vornehmlich in der Corona-Krise und der damit verbundenen Umsatzausfälle im Gastronomiebereich, im Einzelhandel und in der Tourismusbranche rückte der Liquiditätsvorteil von Gutscheinen in den Fokus von Unternehmern, der Öffentlichkeit und des Staates. Die Bundesregierung verabschiedete dahingehend eine freiwillige Gutscheinlösung für vor dem 8. März 2020 erworbene Eintrittskarten im Kultursektor.[37] Hintergrund ist mit Hilfe dieses Beschlusses die Unternehmen vor Liquiditätsschwierigkeiten und drohender Insolvenzen durch die Rückerstattung der Kosten von stornierten Buchungen zu bewahren.[38] Ebenfalls wurde dieser Lösungsansatz zeitversetzt für die Tourismusbranche bei Pauschalreisen beschlossen.[39]

Interessant ist in diesem Zusammenhang die Frage, ob und unter welchen Voraussetzungen ein möglicher Rückerstattungsanspruch gegenüber der Finanzverwaltung im Zusammenhang mit bereits abgeführter Umsatzsteuer der Anzahlungen erfolgen kann. Hier könnte eine weitere eminente Liquiditätsquelle für Unternehmen liegen.

3. Begriffsdefinitionen

3.1. Der Begriff des Gutscheins

3.1.1. Bisherige Auffassung

Eine gesetzliche Legaldefinition war bis zur Umsetzung der Gutschein-Richtlinie nicht im nationalen UStG verankert.[40] Nach der Rechtslage für vor dem 1.1.2019 ausgestellte Gutscheine wurden diese seitens der Finanzverwaltung hinsichtlich ihrer Eigenart differenziert.[41] Hierbei erfolgt eine Abgrenzung zwischen Preisnachlass- oder Rabattgutscheinen, Waren- und Nennwertgutscheinen.[42] Die Oberfinanzdirektion Niedersachsen beschreibt den Nennwertgutschein als "Einkaufs- oder Geschenkgutscheine über einen bestimmten Nennbetrag, z. B. 10,00 EUR, die bei dem ausstellenden Einzelhändler gegen eine beliebige Ware oder Dienstleistung eingetauscht werden können. Hiervon abzugrenzen sind die sog. Warengutscheine, die sich auf eine konkret bezeichnete Ware oder Dienstleistung beziehen (z. B. Kinogutschein)".[43] In der Ausgabe eines Wertgutscheines wurde aus Sicht der Finanzverwaltung bisher lediglich ein Tausch von Zahlungsmitteln (Bargeld in Gutschein) gesehen.[44] Ein Preisnachlass- oder Rabattgutschein ermöglicht dem Kunden einen prozentual oder betragsmäßig niedrigeren gegenüber dem ursprünglichen Preis für eine Ware oder Dienstleistung aufzuwenden.[45]

3.1.2. Gesetzliche Neuregelung zum 1.1.2019

Gutschein i.S.v. § 3 Abs. 13 UStG

Mit Jahressteuergesetz 2018[46] wurde die RL 2016/1965 in das deutsche UStG umgesetzt und § 3 UStG um die Absätze 13 bis 15 erweitert. Die darin enthaltenen Regelungen definieren den Gutschein an sich im umsatzsteuerlichen Sinne und kategorisieren diesen in Einzweck- und Mehrzweckgutscheine.

Gemäß § 3 Abs. 13 UStG wird ein Gutschein (Einzweck- oder Mehrzweck-Gutschein) als Instrument definiert, bei dem die Verpflichtung besteht, es als vollständige oder teilweise Gegenleistung für eine Lieferung oder sonstige Leistung anzunehmen und der Liefergegenstand oder die sonstige Leistung oder die Identität des leistenden Unternehmers auf dem Instrument selbst oder in damit zusammenhängenden Unterlagen, einschließlich der Bedingungen für die Nutzung dieses Instruments, angegeben sind. Die Gutscheine sind als äquivalent zu einer Geldzahlung anzusehen und dementsprechend muss dieser vom Unternehmer auch akzeptiert werden, ansonsten kann kein Gutschein im umsatzsteuerlichen Sinne vorliegen.[47] Folglich können Zahlungsmittel wie Bargeld oder Schecks keine Gutscheine i.S.d. der Definition darstellen.[48] Bezüglich der Ausgabe des Instruments ist keine Formvorschrift gegeben. Gemäß A 3.17 Abs. 1 S. 2 UStAE können diese in physischer (z.B. Papierdokument oder Plastikkarte) als auch in elektronischer Form ausgestellt sein.[49] Nach A 3.17 Abs. 1 S. 3 UStAE ist es unbedenklich für das Vorliegen eines Gutscheins i.S.v. § 3 Abs. 13 UStG, wenn neben dem auf dem Gutschein genannten Nennwert eine Zuzahlung geleistet werden muss, um die Leistung vollständig zu begleichen.

Kern der Beurteilung eines Gutscheins i.S.d. UStG ist die Gewährung eines Rechts auf Erhalt einer Lieferung oder einer

[37] Vgl. „Gesetz zur Abmilderung der Folgen der COVID-19-Pandemie im Veranstaltungsrecht und im Recht der Europäischen Gesellschaft (SE) und der Europäischen Genossenschaft (SCE)".

[38] Vgl. Deutscher Bundestag, „Gesetzentwurf der Fraktionen der CDU/CSU und SPD: Entwurf eines Gesetzes zur Abmilderung der Folgen der COVID-19-Pandemie im Veranstaltungsvertragsrecht." BT-Drs. 19/18697.

[39] Vgl. „Gesetz zur Abmilderung der Folgen der COVID-19-Pandemie im Pauschalreisevertragsrecht und zur Sicherstellung der Funktionsfähigkeit der Kammern im Bereich der Bundesrechtsanwaltsordnung, der Bundesnotarordnung, der Wirtschaftsprüferordnung und des Steuerberatungsgesetztes während der COVID-19-Pandemie".

[40] Vgl. Hans Nieskens, „§ 3 UStG." In Rau/Dürrwächter UStG (Köln: Dr. Otto Schmidt), 189. Lieferung:Rz. 418.

[41] Vgl. Michel, „Internet-Marktplätze und Gutscheine im Fokus der Steuergesetzgebung - umsatzsteuerrechtliche Änderungen zum 01.01.2019", 434; Feil und Polok, „Hält die EU-Gutschein- Richtlinie was sie verspricht?", 65.

[42] Vgl. Rudolf Hahn, „§ 3 UStG." In BeckOK UStG, hrsg. von Weymüller Rainer (München: C.H. Beck oHG), Rz. 601; Hans Nieskens, „§ 3 UStG." In Rau/Dürrwächter UStG, 189. Lieferung:Rz. 418.1.

[43] OFD Niedersachsen, „Verf. v. 17.1.2011 - S 7100-779-St 171." BeckVerw (2011): 246324, 246324.

[44] Vgl. ebd., S.246324; OFD Karlsruhe, „Vfg. v. 13.8.2019 -USt-Kartei S 7270 - Karte 3." BeckVerw (2019): 456763, S. 456763.

[45] Vgl. Michel, „Internet-Marktplätze und Gutscheine im Fokus der Steuergesetzgebung - umsatzsteuerrechtliche Änderungen zum 01.01.2019", 434.

[46] „Gesetz zur Vermeidung von Umsatzsteuerausfällen beim Handel mit Waren im Internet und zur Änderung weiterer steuerlicher Vorschriften." BGBl. I, Nr. 45 (2018): 2338–2353.

[47] Vgl. Daniel Denker und Nicolas Schürkötter, „Umsatzbesteuerung von Gutscheinen in der Praxis: Wieviel Klarheit bringt das BMF-Schreiben vom 2.11.220." USt direkt digital, Nr. 23 (2020): 14-19, S. 15; Ferdinand Huschens, „Einzweck- und Mehrzweck-Gutscheine in der USt - Praktische Auswirkungen der ab 01.01.2019 geltenden Neuregelungen." DB, Nr. 09 (2019): 446–456, S. 450.

[48] Vgl. Denker und Schürkötter, „Umsatzbesteuerung von Gutscheinen in der Praxis", S. 15.

[49] Siehe auch Richtlinie (EU) 2016/ 1065 v. 27.6.2016, Erwägungsgrund 6.

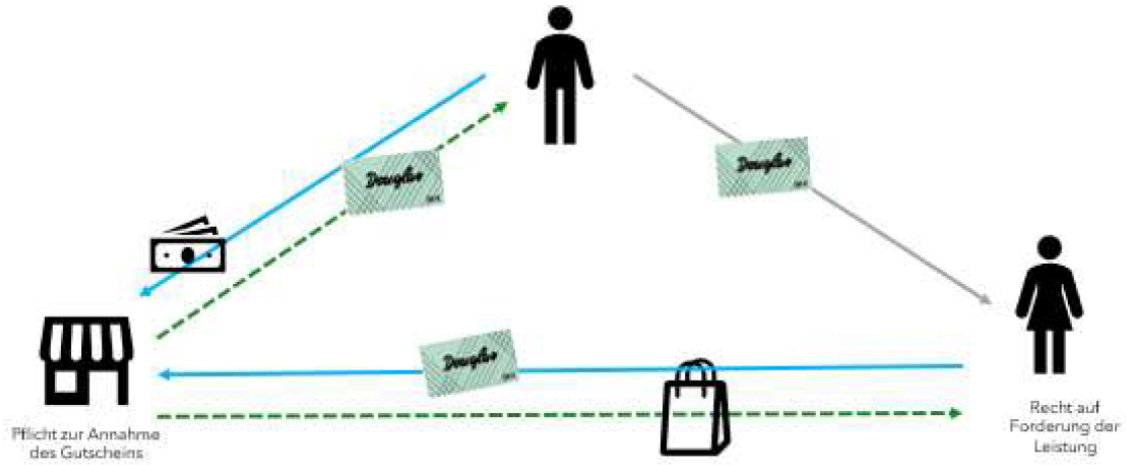

Abbildung 2: Definition eines Gutscheins i.S.v. § 3 Abs. 13 UStG.
In Anlehnung an: Julia Bader et al., Umsatzsteuerrecht visualisiert: Das deutsche Umsatzsteuerrecht in Übersichten, 3., überarbeitete und aktualisierte Auflage (Stuttgart: Schäffer-Poeschel Verlag, 2020), S. 18.

sonstigen Leistung.[50] Daher sind explizit solche Instrumente, die lediglich einen Preisnachlass oder einen Rabatt gewähren und dem Erwerber nicht das Recht auf Erhalt eines Gegenstandes oder einer Dienstleistung gewähren, gemäß § 3 Abs. 13 S. 2 UStG abzugrenzen.[51] Aufgrund fehlenden Entgelts fallen Gutscheine für Warenproben oder Muster nicht unter den genannten Anwendungsbereich. Des Weiteren werden nach A 3.17 Abs. 1 S. 6 UStAE Instrumente wie Fahrscheine, Eintrittskarten für Kinos oder Museen, Briefmarken oder Ähnlichem nicht von § 3 Abs. 13 UStG tangiert.[52] Der BFH konkretisierte in der Vergangenheit, dass „der wesentliche Inhalt des Geschäfts [..] der mit dem Erwerb einer Eintrittskarte verbundene Anspruch auf eine genau bezeichnete Dienstleistung"[53] des genau bezeichneten Unternehmers ist. Die Eintrittskarte stellt demzufolge vielmehr einen Zahlungsnachweis dar.[54] Die Abgrenzung zu einem Gutschein zeichnet sich folglich dadurch aus, dass eine Eintrittskarte inhaltlich genau bestimmt, wohingegen der Leistungsgegenstand beim Gutschein nur bestimmbar, ist.[55] Analog verhält es sich bei Fahr- oder Flugscheinen, da hier der Anspruch auf eine explizit benannte Beförderung besteht.[56]

Aufgeworfene Zweifelsfragen hinsichtlich der Abgrenzung zu bloßen Zahlungsmitteln[57] wurden von der Finanzverwaltung durch A 3.17 Abs. 1 S. 7 UStAE dahingehend geklärt, dass die Remonetarisierbarkeit, d.h. die Möglichkeit einer Rückumwandlung in den ursprünglich gezahlten bzw. den noch nicht verwendeten Betrag, eine Guthabenkarte und damit ein bloßes Zahlungsmittel darstellt. Zahlungsdienste i.S.d. RL 2007/64/EG werden nach A 3.17 Abs. 1 S. 8 UStAE ebenso wenig vom Gutscheinbegriff erfasst.

Der Begriff Einzweck- und Mehrzweck-Gutschein

Gemäß § 3 Abs. 14 UStG hat der Einzweck-Gutschein[58] zunächst die Definitionskriterien aus § 3 Abs. 13 UStG zu erfüllen. Weiter kennzeichnet er sich durch das Vorliegen aller notwendigen Informationen zur Besteuerung aus. Darunter fallen laut Gesetzestext der Ort der Lieferung oder der sonstigen Leistung, auf den sich der Gutschein bezieht, und die für diesen Umsatz geschuldete Steuer im Zeitpunkt der Ausstellung des Gutscheins.

Die Finanzverwaltung erweitert die Definition zur Annahme eines Einzweck-Gutscheins in A 3.17 Abs. 2 UStAE. Hiernach ist – neben den gesetzlich verankerten Kriterien – die Identität des leistenden Unternehmers anzugeben sowie die Leistung insoweit zu konkretisieren, sodass der steuerberechtigte EU-Mitgliedstaat den Steuersatz und dementsprechend den daraus resultierenden Steuerbetrag mit Sicherheit bestimmen kann. Weitere Erfordernis ist, dass zur korrekten Ortsbestimmung bei einer sonstigen Leistung, der Status des Empfängers (Privatkunde oder Unternehmer i.S.d. § 2 UStG) feststehen muss. Bei einem Leistungsgegenstand, zu der der

[50] Vgl. Axel Leonard und Martin Robisch, „§ 3 UStG." In Bunjes UStG, Rn. 323; Frank Fritsch, „§ 3 UStG." In Reiß/Kraeusel/Langer UStG (Bonn: Stollfuß), 164. Lieferung:Rn. 725.2.

[51] Siehe auch Richtlinie (EU) 2016/ 1065 v. 27.6.2016, Erwägungsgrund 4.

[52] Siehe auch ebd., Erwägungsgrund 5.

[53] BFH, „Urt. v. 25.4.2018 - XI R 16/16 mAnm Müller." MwStR (2018): 880-884, S. 882, Rn. 26; BFH, „Urt. v. 3.6.2009 - XI R 34/08." DStR (2009): 2369-2371, S. 2370.

[54] Vgl. Denker und Schürkötter, „Umsatzbesteuerung von Gutscheinen in der Praxis", S. 15; Ferdinand Huschens, „Einzweck- und Mehrzweck-Gutscheine - Die zu erwartende Rechtsauffassung der Verwaltung." UVR (2020): 48–56, S. 49.

[55] Vgl. Suse Martin, „§ 3 UStG." In Sölch/Ringleb Umsatzsteuergesetz, 89. EL Juni 2020:Rn. 793.

[56] EuGH, „Urt. v. 23.12.2015 - C-250/14, C-289/14 - Air France-KLM, Hop!-Brit Air SAS." DStR (2016): 118-122, S. 120-121, Rn. 28.

[57] Vgl. Deutscher Industrie- und Handeslkammertag e.V. et al., „Stellungnahme zum Umsetzungsvorschlag der Gutschein-Richtline (EU) 2016/1065 durch die Finanzverwaltung", S. 7.

[58] Im Englischen wird der Einzweck-Gutschein als Single Purpose Voucher (SPV) bezeichnet.

Gutschein berechtigt, ist dieser zumindest hinsichtlich seiner Gattung anzugeben. Letztendlich soll der Einzweck- Gutschein vom Aussteller sichtbar als solcher gekennzeichnet werden.

Gemäß der in § 3 Abs. 15 UStG enthaltenen Legaldefinition liegt dann ein Mehrzweck- Gutschein[59] vor, wenn er die Merkmale aus § 3 Abs. 13 UStG erfüllt, jedoch kein Einzweck- Gutschein ist. Ergänzt wird diese Definition in A 3.17 Abs. 9 S. 4 UStAE dahingehend, dass es sich insbesondere um einen Mehrzweck-Gutschein handelt, wenn dieser für Leistungen zum ermäßigten oder regulären Steuersatz eingetauscht werden kann. Parallel zur Kennzeichnung des Einzweck-Gutscheins als solcher, hat dies gemäß A 3.17 Abs. 9 S. 6 UStAE auch bei Mehrzweck-Gutscheinen zu erfolgen.

Anwendung

Wie oben bereits erwähnt gelten die gesetzlichen Regelungen seit dem 1.1.2019, jedoch hat die Finanzverwaltung im Wege der Billigkeit in seinem BFM-Schreiben vom 2.11.2020 eine großzügige Nichtanwendungsregelung gewährt. Demzufolge wird es nicht bemängelt, wenn die ab dem 1.1.2019 und vor dem 2.2.2021 ausgestellte Gutscheine nicht den Vorgaben der Finanzverwaltung entsprechen. Hierzu ist zu erwähnen, dass Verwaltungsanweisungen wie das BMF-Schreiben und der Umsatzsteuer-Anwendungserlass grundsätzlich nur die Finanzverwaltung, jedoch nicht die Finanzgerichte und auch nicht den Steuerpflichtigen, binden.[60]

Kritische Analyse

Da die Klassifizierung des Gutscheins erhebliche Auswirkungen auf die Besteuerung hat, soll diese im ersten Schritt näher beleuchtet werden. Zunächst ist festzuhalten, dass der deutsche Gesetzgeber die Richtlinie im Wortlaut weitestgehend identisch umgesetzt hat. Auffallend ist jedoch, dass das UStG um den Hinweis auf die negative Abgrenzung des Preisnachlasses, ergänzt wurde. Diese Ergänzung ist m.E. nicht essenziell, da diese bereits durch die Definition in § 3 Abs. 13 UStG ersichtlich wird. Schließlich wurde die Abgrenzung hinsichtlich der Eintrittskarten auch nicht im Gesetz verankert.

Ein Vergleich der Begriffsdefinitionen nach alter und neuer Rechtslage lässt zunächst den Eindruck erwecken, dass der Einzweck-Gutschein durchaus mit dem nach alter Rechtslage definiertem Warengutschein und der Mehrzweck-Gutschein mit dem Nennwertgutschein vergleichbar ist.[61]

Neue Einordnung des Nennwertgutscheins

Es scheint, dass zur Annahme eines Einzweck-Gutscheins ein enger Rahmen durch das Gesetz und das BMF gesetzt wurde. Jedoch ist es nun regelmäßig der Fall, dass die im Einzelhandel repräsentativen Nennwertgutscheine unter die Definition der Einzweck- Gutscheine fallen. Darunter dürften eine große Anzahl der ausgegebenen Gutscheine lokaler Einzelhändler fallen, wo exklusiv der reguläre (oder der ermäßigte) Steuersatz zur Anwendung kommt und der Ort der Leistung im Inland ist.

Beispiel:

In einem Möbelhaus in Landsberg am Lech erwirbt ein Kunde einen klassischen Nennwert- Gutschein über einen Betrag von 200 EUR für 200 EUR. In diesem Möbelhaus ist ausschließlich der Erwerb von Waren, die dem Regelsteuersatz unterliegen, möglich. Da sowohl Ort als auch die Steuerschuld bereits feststehen, liegen die Voraussetzungen nach § 3 Abs. 14 UStG für einen Einzweck-Gutschein vor.[62]

Eine Vielzahl von Unternehmen haben typischerweise Nennwertgutscheine an ihre Kunden ausgegeben, wobei vor dem 1.1.2019 lediglich von einem Tausch von Zahlungsmitteln ausgegangen wurde. Eine dementsprechende Umqualifizierung hat ein wesentliches Anpassungspotential in Bezug auf das operative Geschäft sowie die dahinterstehenden Prozesse. Von einer Umqualifizierung sind zahlreiche Abteilungen betroffen. Wie bereits angesprochen leisten Gutscheine einen wesentlichen Liquiditätsbeitrag. Dementsprechend ist durch das Controlling bei einer Besteuerung im Zeitpunkt der Ausgabe von Gutscheinen Liquiditätsabfluss an das Finanzamt zu berücksichtigen. Demensprechend sind auch die internen Systeme (ERP, Kasse, etc.) in Bezug auf die richtige umsatzsteuerliche Würdigung zu prüfen und anzupassen. Weiter sind interne Prozesse zu aktualisieren und die beteiligten Mitarbeiter im Rechnungswesen, Controlling und Vertrieb hinsichtlich der Umqualifizierung und der daraus ergebenden umsatzsteuerlichen Fragestellungen und Auswirkungen zu schulen. Da der Gutschein ein wichtiges Absatzinstrument darstellt, ist die Marketingabteilung insbesondere bezüglich der umsatzsteuerrechtlichen Konsequenzen im Hinblick auf die Vielzahl der Gutscheine zu sensibilisieren.

Identität des leistenden Unternehmers

Die in § 3 Abs. 13 Nr. 2 UStG zweite Voraussetzung zur Bestätigung eines Gutscheins besagt, dass entweder die Leistung oder die Identität des leistenden Unternehmers anzugeben sind. Eine dementsprechende gesetzliche Pflicht zur Klassifizierung eines Einzweck-Gutscheins ist gesetzlich nicht verankert, wird nun aber durch das BMF in A 3.17 Abs. 2 S. 2 UStAE gefordert, wodurch ein Widerspruch zum Gesetz

[59] Im Englischen wird der Mehrzweck-Gutschein als Multi Purpose Voucher (MPV) bezeichnet.

[60] Vgl. Ulrich Krömker, „§ 1 FGO." In Basiskommentar Steuerrecht, hrsg. von Otto-Gerd Lippross und Wolfgang Seibel. 88. Ergänzungslieferung (Dr. Otto Schmidt), 116. Ergänzungslieferung:Rn. 3.

[61] Vgl. Michel, „Internet-Marktplätze und Gutscheine im Fokus der Steuergesetzgebung - umsatzsteuerrechtliche Änderungen zum 01.01.2019", S. 436.

[62] In Anlehnung an: Heinrichshofen und Kupke, „Gutscheine nach der Änderung der Mehrwertsteuer- Systemrichtlinie (MwStSystRL) - Anmerkungen zu den Auswirkungen auf die deutsche Besteuerungspraxis", S. 147.

erkennbar ist. Weiter wird nicht konkretisiert, was mit der Identität des Unternehmers exakt gemeint ist.

Dies könnte in der Praxis dahingehend Problematiken aufweisen, dass im Einzelhandel oftmals einzelne Unternehmer über die Zugehörigkeit zu einem Unternehmernetzwerk zusammengeschlossen sind. Beispielhaft sei hier die Modemarke Esprit zu nennen. Esprit vertreibt seine Waren über einen Online-Shop (Esprit Retail B.V. & Co. KG) als auch eigene Stores (Esprit Store Augsburg) sowie Franchiseunternehmer (Esprit Franchise Store Mindelheim).

Nach oben genannten Kriterien ist der Geschenkgutscheins als Einzweck-Gutschein einzuordnen, da zum einen die Definition als Gutschein i.S.v. § 3 Abs. 13 vorliegt und zum anderen liegt der Ort der Lieferung, unter der Annahme, dass der Gutschein nur innerhalb von Deutschland einlösbar, fest. Weiter unterliegt der Leistungsgegenstand (Modeartikel) exklusiv dem regulären Steuersatz. Nun stellt sich jedoch die konträre Notwendigkeit seitens des Gesetzgebers und der Finanzverwaltung zur Angabe der Identität des leistenden Unternehmers. Auch wenn der Store in Augsburg als Betriebsstätte nach § 12 AO anzusehen ist, so ist der Franchisenehmer ein eigenständiger Unternehmer.[63] Die Gutscheine sind regelmäßig im Onlineshop als auch in allen teilnehmenden Geschäften einzulösen. Der genannte Franchise Store Mindelheim akzeptiert die Geschenkgutscheine als Zahlungsmittel.[64]

Es stellt sich nun die Frage, ob und inwieweit die Identität des leistenden Unternehmers anzugeben ist. Seitens der Finanzverwaltung wird nicht konkretisiert, ob die Identität direkt auf dem Gutschein aufgedruckt zu sein hat oder diese sich auch aus einer Vielzahl anderer Dokumente (Homepage, Verweise auf AGB, etc.) ergeben kann. Folglich besteht erhebliche Rechtsunsicherheit hinsichtlich der Folgen bei fehlenden Angaben. M.E. ist nach Gesetzestext der Gutschein weiterhin als Einzweck-Gutschein zu klassifizieren. Dagegen geht Huschens bei einer fehlenden Identität des leistenden Unternehmers regelmäßig von einem Mehrzweck-Gutschein aus.[65] Eine dementsprechende Bitte um Präzisierung wurde bereits vor Veröffentlichung des BMF-Schreibens an die Finanzverwaltung gerichtet.[66]

Angabe der Gattung

Die Finanzverwaltung fordert nach A. 3.17 Abs. 2 S. 4 UStAE die Angabe der jeweiligen Gattung des Leistungsgegenstands auf dem Einzweck-Gutschein. Als Gattung ist die Gesamtheit von Arten von Waren oder sonstigen Leistungen zu verstehen, die durch ihre wesentliche Eigenschaft insoweit übereinstimmen, um den Steuersatz eindeutig zu bestimmen. Die Finanzverwaltung spricht in seinen Ausführungen von einer Pflichtangabe („muss"). Im Umkehrschluss heißt das, dass eine fehlende Gattungsbezeichnung dazu führt, dass der Gutschein nicht als Einzweck-Gutschein qualifiziert wird. Dabei stellt sich die Frage, ob dies auch dann gilt, wenn das Warensortiment oder die Dienstleistung des leistenden Unternehmers insgesamt demselben Steuersatz unterliegt. Meiner Meinung nach ist keine Änderung hinsichtlich der umsatzsteuerlichen Würdigung zu erkennen, wenn ein Modeunternehmen auf ihrem Gutschein keine Gattungsbezeichnung oder die Bezeichnung „Warensortiment" oder die Gattungsbezeichnung „Kleidung und Accessoires" verwendet. In der Literatur scheint hierzu kein übereinstimmender Konsens zu herrschen.[67]

Zudem ist hier ein Widerspruch zum Gesetz erkennbar. Gemäß § 3 Abs. 13 Nr. 2 UStG ist es ein Wahlrecht gegeben, entweder die Leistung oder die oben dargestellte Identität des leistenden Unternehmers auf dem Gutschein selbst bzw. den damit zusammenhängenden Unterlagen anzugeben.

Kennzeichnungspflicht

Fraglich erscheint weiter die vom BMF geforderte Kennzeichnungspflicht des jeweiligen Gutscheins. Die Forderung wird als „Soll"-Vorschrift formuliert, daher ergibt sich hier durchaus eine gewisse Rechtsunsicherheit, insoweit auf diese Kennzeichnung verzichtet wird. M.E. kann die Kennzeichnung kein Definitionskriterium für einen Gutschein als Einzweck- oder Mehrzweck-Gutschein darstellen. Die entscheidenden Definitionskriterien sind im Gesetz verankert, das BMF-Schreiben gibt hierzu eine Hilfestellung und ist dem Grunde nach nur für die Finanzverwaltung bindend. Zudem ist der zugrundeliegenden EU- Richtlinie hierzu keine Regelung zu entnehmen. Daher ist eine Sanktionierung bei Verstoß eher als nicht durchsetzbar anzusehen.

Jedoch ist aufgrund der Bindungskraft der Finanzverwaltung an die Verwaltungsanweisung Diskussionspotential im Rahmen einer Betriebsprüfung gegeben. Die Finanzverwaltung wird ihre Auffassung und damit die Kennzeichnungspflicht gemäß der Verwaltungsanweisung vertreten. Aus Sicht des Unternehmens sind dabei zwei Szenarien möglich. Entweder ist – trotz fehlender Kennzeichnung – die Kategorisierung des Gutscheins und der daraus erfolgten umsatzsteuerlichen Würdigung korrekt oder fehlerhafte erfolgt. Beide Szenarien sind jedoch nach meinem Dafürhalten

[63] Vgl. Gottfried Löwisch, „§ 84 HGB." In Handelsgestzbuch Band 1, hrsg. von Carsten Thomas Ebenroth et al. 4. Auflage (C.H. Beck/Verlag Vahlen, 2020), Rn. 208.

[64] Vgl. Esprit Global Image GmbH, „Esprit Fanchise Store Mindelheim.". https://www.esprit.com/storefinder?storeid=13470872&location=48.0459,10.4872 (letzter Zugriff: 18. März 2021).

[65] Vgl. Huschens, „Einzweck- und Mehrzweck-Gutscheine - Die zu erwartende Rechtsauffassung der Verwaltung", S. 50.

[66] Vgl. Deutscher Industrie- und Handeslkammertag e.V. et al., „Stellungnahme zum Entwurf eines BMF- Schreibens zur umsatzsteuerlichen Behandlung von Einzweck- und Mehrzweck-Gutscheinen: GZ: III C 2 - S7100/19/10001:002; DOK 2019/1090742.". https://www.gdv.de/resource/blob/56412/d42d59ae582f10685246106db38d40d0/stellungnahme-gutscheine---download-data.pdf (letzter Zugriff: 5. März 2021), S. 1.

[67] Vgl. Stefanie Becker, „Behandlung von Gutscheinen im Umsatzsteuerrecht: Offene Fragen nach dem BMF- Schreiben v. 2.11.2020 sowie Auswirkungen für die Praxis." MwStR (2021): 223–229, S. 224; Huschens, „Einzweck- und Mehrzweck-Gutscheine - Die zu erwartende Rechtsauffassung der Verwaltung", S. 50.

nicht auf die Kennzeichnung zurückzuführen.

Es scheint, dass diese Bedingung lediglich als eine Art Vertrauensschutz für die innerhalb von komplexen Vertriebsketten agierenden Unternehmer darstellen soll, was jedoch nach A 3.17 Abs. 2 S. 9 UStAE keinesfalls von einer gewissen Prüfungspflicht befreit. Ist der Aussteller des Gutscheins zugleich auch Leistender, ist eine Kennzeichnung in meinen Augen obsolet und führt bei einer Vielzahl von klein- und mittelständischen Unternehmen im Einzelhandel zu einem zusätzlichen Kostenfaktor, da in der Regel Gutscheinformate/Gutscheingeschenkkarten bereits in einer großen Anzahl bestellt wurden und vorgedruckt gelagert werden. Hier sollte aus praxis- und kostenrelevanten Gründen auf eine Anpassung verzichtet werden, da keine Gefährdung des Steueraufkommens für den Staat gegeben ist.

Ist der Gutscheinaussteller nicht auch der leistende Unternehmer, ist eine Präzisierung bzgl. der Qualifikation des Gutscheins für die folgenden Unternehmer in der Vertriebskette als auch für die Betriebsprüfung wünschens- und auch empfehlenswert.[68] Jedoch sind diese Feinheiten meiner Ansicht auch als Teil der Kaufverträge niederzuschreiben. In der Regel sollte davon auszugehen sein, dass bei Vermittlungsleistungen eine Vielzahl von Gutscheinen gehandelt wird und dementsprechende Kaufverträge aufgesetzt sind. Es dürfte eher unüblich sein, dass lediglich einzelne Gutscheine innerhalb von Vertriebsketten verkauft werden.

Zusammenfassend kann festgehalten werden, dass die deutsche Finanzverwaltung zum Gesetz abweichende Anforderungen an die Qualifizierung als Einzweck-Gutschein stellt. Es erscheint mir nicht dem Ziel gerecht zu werden eine einheitliche unionsweite Lösung zu etablieren. Mit dem BMF-Schreiben sind Hilfestellungen hinsichtlich der Definition eines Gutscheins begrüßenswert, jedoch werden m.E. mit den zusätzlichen Anforderungen neuer Abstimmungsbedarf und Rechtsunsicherheiten geschaffen.

3.2. Weitere Begriffsdefinitionen aus den Steuergesetzen

3.2.1. Begriff der ruhenden Lieferung

Gemäß § 3 Abs. 7 S. 1 UStG liegt eine ruhende Lieferung dann vor, wenn der Gegenstand der Lieferung nicht befördert oder versendet wird. Die Lieferung gilt dann als dort ausgeführt, wo sich der Gegenstand zur Zeit der Verschaffung der Verfügungsmacht befindet. Um die Verschaffung der Verfügungsmacht positiv zu bestätigen sind zwei Elemente erforderlich. Zum einen muss der Substanzwert übertragen werden und zum anderen muss der Wille des Leistenden den Substanzwert an den Leistungsempfänger endgültig zu übertragen.[69] Charakteristisch für ruhende bzw. unbewegte Lieferungen sind Grundstücke.[70] Nicht beachtlich bei der Bestimmung des Leistungsortes sind beispielsweise der Ort des Vertragsabschlusses des zugrunde liegenden Umsatzgeschäftes oder die Nationalität und Aufenthaltsort des leistenden Unternehmers und des Leistungsempfängers.[71] In der Regel hat die Ortsbestimmung in den Fällen des § 3 Abs. 7 S. 1 UStG geographisch punktbezogen zu erfolgen.[72]

3.2.2. Begriff der Anzahlung

Der Anzahlung liegt ein vom Gesetzgeber normierter Sondertatbestand der Ist- Besteuerung in § 13 Abs. 1 Nr. 1 Buchst. a S. 4 UStG zugrunde. Hiernach entsteht die Steuer mit Ablauf des Voranmeldezeitraums, in dem das Entgelt oder das Teilentgelt vereinnahmt wurde. Eine Vorauszahlung qualifiziert sich dahingehend, dass der gesamte Rechnungsbetrag vor Leistungsausführung oder Teilleistungsausführung abgerechnet wird. Dagegen wird bei Abrechnung eines Teilbetrages des vereinbarten Rechnungsbetrags von einer Anzahlung oder Abschlagszahlung gesprochen.[73] Die Vorverlagerung kann nur so lange Bestand haben, wie der Steueranspruch nicht wegen einer Leistungsausführung entstanden ist.[74]

Voraussetzung für eine vorgezogene Besteuerung ist zum einen die Vereinnahmung des Entgelts vor der Leistungsausführung. Eine Vereinnahmung ist zu bejahen, wenn der leistende Unternehmer wirtschaftlich verfügen kann.[75] Zum anderen hat die Leistung hinsichtlich Art, Umfang und Zeitpunkt vor dem Hintergrund der Anwendbarkeit von Steuerbefreiungen oder des Steuersatzes bestimmbar zu sein.[76]

3.2.3. Begriff der Vermittlungsleistung

Eine Definition der Vermittlungsleistung ist im Gesetz nicht etabliert. Inhalt einer Vermittlungsleistung ist vom Vermittler das Erforderliche zu tun, damit zwei Parteien einen Vertrag schließen, wobei er selbst kein Interesse am Inhalt des Vertrages hat.[77] Weiter qualifiziert sie sich dadurch, dass sie einer Partei erbracht wird und von dieser als Vermittlungstätigkeit (Dienstleistung) vergütet wird.[78] Zivilrechtlich kommt das vermittelte Geschäft zwischen Auftraggeber und dem Dritten zustande, somit ist der Vermittler in die Leistungserbringung nicht eingebunden.[79] Dabei erfordert die Vermittlungsleistung ein Handeln in fremden Namen und für fremde Rechnung.[80] Ein dementsprechender Ausdruck des Willen hat gegenüber dem Beteiligten deutlich zum Ausdruck zu kommen.[81] A 3.7 Abs. 1 S. 3 UStAE besagt, dass eine

[68] Vgl. Robert C. Prätzler, „BMF nimmt Stellung zu umsatzsteuerlichen Fragen bei Gutscheinen." DB, Nr. 47 (2020): 2491-2493, S. 2492.

[69] Vgl. Frank Fritsch, „§ 3 UStG." In Reiß/Kraeusel/Langer UStG (Bonn: Stollfuß), 165. Lieferung:Rn. 132.

[70] Vgl. Suse Martin, „§ 3 UStG." In Sölch/Ringleb Umsatzsteuergesetz (München: C.H. Beck oHG), 90. EL:Rn. 504.

[71] Vgl. Hans Nieskens, „§ 3 UStG." In Rau/Dürrwächter UStG (Köln: Dr. Otto Schmidt), 191. Lieferung:Rn. 3607.

[72] Ebd., 191. Lieferung:Rn. 3609.

[73] Vgl. Hans Nieskens, „§ 13 UStG." In Rau/Dürrwächter UStG (Köln: Dr. Otto Schmidt), 183. Lieferung:Rz. 321.

[74] Vgl. ebd., 183. Lieferung:Rn. 296.

[75] Vgl. ebd., 183. Lieferung:Rn. 326.

[76] Vgl. Carsten Leipold, „§ 13 UStG." In Sölch/Ringleb Umsatzsteuergesetz (München: C.H. Beck oHG), Rn. 88.

[77] Vgl. EuGH, „Urt. v. 21.6.2007 - C-453/05 - Volker Ludwig." DStR (2007): 1160-1164, Rn. 23.

[78] Vgl. Suse Martin, „§ 3 UStG." In Sölch/Ringleb Umsatzsteuergesetz, 90. EL:Rn. 552.

[79] Vgl. EuGH, „Urt. v. 14.7.2011 - C-464/10 - Pierre Henfling." IStR (2011): 593-595, Rn. 33.

[80] Vgl. BFH, „Urt. v. 18.05.1994 - XI R 62/93." BeckRS (1994): 22011100.

[81] Vgl. BFH, „Urt. v. 19.01.1967 - V 52/63." BeckRS (1967): 21005116.

Vermittlungsleitung umsatzsteuerrechtlich entsprechend der Regelung des § 164 Abs. 1 BGB grundsätzlich nur vorliegt, insoweit der Vermittler das Umsatzgeschäft erkennbar im Namen des Vertretenen abgeschlossenen hat.

Bei der Ortsbestimmung der Vermittlungsleistung ist auf den Empfänger der Leistung abzustellen. Bei Vermittlungsleistungen an einen Unternehmer bestimmt sich der Ort nach den allgemeinen Regelungen nach § 3a Abs. 2 UStG. Demnach gilt der Umsatz als dort ausgeführt, von wo aus der Empfänger sein Unternehmen betreibt. § 3a Abs. 3 Nr. 4 UStG regelt den Ort einer Vermittlungsleistung für nichtunternehmerische Zwecke. Die Vermittlungsleistung wird dort erbracht, an dem der vermittelte Umsatz als ausgeführt gilt.

3.2.4. Begriff der missbräuchlichen rechtlichen Gestaltungsmöglichkeit

§ 42 AO definiert eine Aufforderung Umgehungen der Steuergesetze, soweit sie auf einem Missbrauch von Gestaltungsmöglichkeiten des Rechts beruhen, zu erkennen und die genannten Rechtsfolgen anzuwenden. § 42 Abs. 2 AO beinhaltet eine gesetzliche Missbrauchsdefinition.

Gemäß § 38 AO entsteht der Steueranspruch nur, wenn der Tatbestand des jeweiligen Gesetzes verwirklicht ist. Folglich entsteht kein Steueranspruch, wenn der Tatbestand auch nicht verwirklich ist. Regelmäßig werden von Steuerpflichtigen ihre unter den gesetzlichen Tatbestand untergestellten Sachverhalt gezielt so gestaltet, dass dieser Tatbestand nicht zur Anwendung kommt.[82] Daher ist in jedem Fall eine zweckgerichtete Handlung erforderlich.[83] Die Annahme einer Steuervermeidung statt einer -umgehung liegt vor, wenn ein Sachverhalt verwirklicht wird, welcher keine Besteuerung bewirkt. An sich ist eine Steuergestaltung und -umgehung weder verboten noch strafbar.[84] Der Steuerpflichtige ist dahingehend in seiner Gestaltung frei die Gegebenheiten so zu konzipieren, dass eine für ihn eine geringe oder keine Steuerbelastung entsteht.[85] Hiervon abzugrenzen ist eine Steuerhinterziehung nach AEAO zu § 42 Nr. 3 insoweit, da hier der Steuerpflichtige unvollständige oder unrichtige Angaben gegenüber dem Finanzamt macht, um das Vorliegen einer Steuerumgehung zu verschleiern.

Zentraler Prüfpunkt des § 42 AO ist der Begriff des Missbrauchs. Erst bei positivem Vorliegen eines Missbrauchs treten die Rechtsfolgen nach § 42 Abs. 1 S. 3 AO ein. Ein Missbrauch liegt nach § 42 Abs. 2 AO vor, wenn eine unangemessene rechtliche Gestaltung zu einem gesetzlich nicht vorgesehenen Steuervorteil führt und der Steuerpflichtige keine beachtlichen Gründe für die Gestaltung nachweisen kann.

[82]Vgl. Klaus-Dieter Drüen, „§ 42 AO." In Tipke/Kruse, AO/FGO (Köln: Verlag Dr. Otto Schmidt), 164. Lieferung 02.2021:Rn. 2.

[83]Vgl. BFH, „Urt. v. 18.3.2004 - III R 25/02." DStR 2004: 1078-1086, S. 1082.

[84]Vgl. Eckart Ratschow, „§ 42 AO." In Klein AO (München: C.H. Beck oHG, 2020), 15. Auflage 2020:Rn. 15.

[85]Vgl. Klaus-Dieter Drüen, „§ 42 AO." In Tipke/Kruse, AO/FGO, 164. Lieferung 02.2021:Rn. 3; Eckart Ratschow, „§ 42 AO." In Klein AO, 15. Auflage 2020:Rn. 42.

4. Umsatzsteuerliche Behandlung von Gutscheinen

4.1. Bisherige deutsche Auffassung

Wie bereits oben erwähnt wurde bisher aufgrund der bislang erlassenen OFD- Verfügungen[86] in Deutschland zwischen (Nenn-)Wertgutscheinen und Waren- oder Sachgutscheinen unterschieden. In der Ausgabe der Nennwertgutscheine wurde bisher aufgrund der Annahme des Zahlungsmitteltauschs kein steuerbarer Vorgang angenommen.[87] Diese Ansicht wurde seitens der Finanzverwaltung[88], der Rechtsprechung[89] als auch in der Literaturmeinung[90] einstimmig vertreten. Eine Anzahlung gem. § 13 Abs. 1 Nr. 1 Buchst. a S. 4 UStG ist aufgrund der fehlenden konkret bezeichneten Leistung zu verneinen.[91] Infolgedessen stellt allein die Hingabe eines solchen Gutscheins kein Leistungsaustausch im umsatzsteuerlichen Sinne dar. Im Zeitpunkt der Einlösung des Wertgutscheins, sprich bei Ausführung des konkreten Umsatzes, entsteht die Umsatzsteuer.[92]

Zwar wird die Leistung bei der Ausgabe eines Warengutscheines – wie beim Wertgutschein ebenfalls noch nicht erbracht, jedoch ist nach Auffassung der Finanzverwaltung, die im Gutschein bezeichnete Leistung bekannt. Ebendaher wird der gezahlte Betrag für den Erwerb dieses Warengutscheins als Anzahlung nach § 13 Abs. 1 Nr. 1 Buchst. a S. 4 UStG auf die genannte Leistung angesehen. Die Anzahlung wiederum unterliegt nach § 13 Abs. 1 Nr. 1 Buchst. a S. 4 UStG der Ist-Besteuerung, d.h. die Rechnung über den Verkauf des Gutscheins ist mit Ausweis deutscher Umsatzsteuer auszustellen.[93]

Auch wenn die OFD-Verfügungen für einfache Fälle in der Praxis gewisse Lösungsansätze bieten, so wurde eine gesetzliche und einheitliche europäische Lösung in der Literatur sehr begrüßt.[94] Wie bereits dargestellt hat sich der Gutscheinmarkt an sich in den letzten Jahren stark gewandelt und

[86]Vgl. OFD Magdeburg, „Verf. v. 2.5.2006 - S 7200 - 179 - St 244." BeckVerw (2006): 91065; OFD Karlsruhe, „Verf. v. 25.8.2011 - S 7270." DStR (2011): 1910; OFD Niedersachsen, „Verf. v. 17.1.2011 - S 7100-779-St 171"; OFD Frankfurt/M, „Verf. v. 25.3.2010 - S 7100 A - 172 - St 110." BeckVerw (2010): 238365.

[87]Vgl. OFD Niedersachsen, „Verf. v. 17.1.2011 - S 7100-779-St 171", S. 246324.

[88]Vgl. OFD Magdeburg, „Verf. v. 2.5.2006 - S 7200 - 179 - St 244"; OFD Niedersachsen, „Verf. v. 17.1.2011 - S 7100-779-St 171".

[89]Vgl. BFH, „Urt. v. 24.8.2006 - V R 16/05." DStR (2006): 2249-2253.

[90]Vgl. Hans Nieskens, „Verkauf von Warengutscheinen: -EuGH-Entscheidung in Sachen Astra Zeneca und ihre Folgen für das deutsche Recht." UR (2010): 893–902, 894; Ulrich Grünwald und Annette Pogodd, „Ausgabe von Wertgutscheinen durch Händler: Hinweise zur Vermeidung umsatzsteuerlicher Risiken." ustb (2001): 212–215, S. 212.

[91]Vgl. Heinrichshofen und Kupke, „Gutscheine nach der Änderung der Mehrwertsteuer-Systemrichtlinie (MwStSystRL) - Anmerkungen zu den Auswirkungen auf die deutsche Besteuerungspraxis", S. 143.

[92]Vgl. OFD Magdeburg, „Verf. v. 2.5.2006 - S 7200 - 179 - St 244".

[93]Vgl. ebd.

[94]Vgl. Feil und Polok, „Hält die EU-Gutschein-Richtlinie was sie verspricht?", S. 68; Feil, Kupke und Greisl, „Aktuelle Entwicklungen bei der umsatzsteuerlichen Behandlung von Gutscheinen", S. 3122; Heinrichshofen und Kupke, „Gutscheine nach der Änderung der Mehrwertsteuer-Systemrichtlinie (MwStSystRL) - Anmerkungen zu den Auswirkungen auf die deutsche Besteuerungspraxis", S. 144.

ausgebaut, wobei die umsatzsteuerliche Würdigung durchaus mit einer großen Rechtsunsicherheit belastet war. Eine entsprechend falsche Würdigung kann für Unternehmer eine immense Steuernachzahlung und schwerwiegende finanzielle Einbußen mit sich bringen. In der Vergangenheit wurde daher oftmals ein komplexeres Geschäftsmodell mit der Möglichkeit einer verbindlichen Auskunft nach § 89 AO abgesichert.[95]

4.2. Rechtslage ab 1.1.2019

4.2.1. Besteuerung außerhalb von Vertriebsketten

Umsatzsteuerliche Würdigung von Einzweck-Gutscheinen

Liegt bei Ausgabe eines Gutscheins ein Einzweck-Gutschein vor, so ist zunächst gemäß § 3 Abs. 14 S. 2 bis 3 UStG zwischen der Übertragung im eigenen Namen oder von einem Unternehmer im Namen eines anderen Unternehmers zu unterscheiden. Überträgt ein Unternehmer einen Einzweck-Gutschein im eigenen Namen, so wird nach § 3 Abs. 14 S. 2 UStG eine Lieferung oder sonstigen Leistung fingiert. Die tatsächliche Leistungserbringung wird gemäß § 3 Abs. 14 S. 5 UStG dann nicht als unabhängiger Umsatz gewürdigt. In Abhängigkeit der Leistung bestimmt sich dementsprechend der Leistungsort. Liegt dem Gutschein eine Lieferung zu Grunde, so bestimmt sich der Ort der fiktiven Lieferung nach Auffassung der Finanzverwaltung nach § 3 Abs. 7 S. 1 UStG, da im Zeitpunkt der Ausgabe keine Waren befördert oder versendet werden. Hierzu vertieft das BMF seine Anmerkungen dahingehend, dass aufgrund der fehlenden Warenbewegung keine Steuerbefreiung für Ausfuhrlieferungen und innergemeinschaftlichen Lieferungen Anwendung findet. Im Falle einer sonstigen Leistung sind die allgemeinen Ortsbestimmungen heranzuziehen, welche regelmäßig vom Status des Leistungsempfängers abhängig sind. Daher verlangt die Finanzverwaltung in A 3.17 Abs. 5 S. 3 UStAE dahingehend bereits bei der Definition als Einzweck- oder Mehrzweck-Gutschein eine Einordnung. Zur Bemessungsgrundlage und dem Steuersatz sind keine eigenen Regelungen genannt. Gemäß A 3.17 Abs. 6 S. 1 UStAE gelten die allgemeinen Regelungen des § 10 Abs. 1 UStG. Ferner ist anzumerken, dass eine besondere Vorsicht bei sonstigen Leistungen geboten ist, wenn der Leistungsempfänger zwar ein Unternehmer ist, jedoch die Leistung ihrer Art nach mit hoher Wahrscheinlichkeit für den privaten Gebrauch bestimmt ist (vgl. A 3a.2 Abs. 11a UStAE).

Gesetzlich ist grundsätzlich nicht geregelt, welche Folgen sich aus der Nichteinlösung von Gutscheinen ergeben. Werden Einzweck-Gutscheine vom Gutscheininhaber nicht eingelöst, ergeben sich nach Verwaltungsauffassung daraus keine weiteren umsatzsteuerlichen Folgen, da die Leistung aufgrund der Fiktion im Zeitpunkt der Ausgabe des Gutscheins bereits als erbracht gilt und der Besteuerung unterworfen wurde. Hier entsteht zwar für den Verkäufer ein Liquiditätsnachteil im Vergleich zu nichteingelösten Mehrzweck-Gutscheinen, dennoch entspricht dies der im Zusammenhang mit Anzahlungen dargelegte Meinung des BFH[96] und EuGH[97]. Eine Änderung der Bemessungsgrundlage nach § 17 UStG ist daher nicht einschlägig. Aus Sicht der Literatur wird diese Behandlung vom Gesetzgeber wohlwissend und bewusst hinsichtlich der dahinterstehenden Fiktion bei Einzweck-Gutscheinen hingenommen.[98]

Umsatzsteuerliche Würdigung von Mehrzweck-Gutscheinen

§ 3 Abs. 15 UStG regelt, dass erst die tatsächliche Erbringung der Lieferung oder sonstigen Leistung bei Einlösen des Mehrzweck-Gutscheins der Umsatzbesteuerung unterliegt. Der Ort der Lieferung oder der sonstigen Leistung bestimmt sich bei Einlösung nach den allgemeinen Anwendungsvorschriften. Wird der Gutschein bei Einlösung als Tausch von Zahlungsmitteln für eine Lieferung benutzt, so richtet sich der Ort der Lieferung nach § 3 Abs. 5a UStG. Wird eine sonstige Leistung als Gegenleistung für den Gutschein angenommen, bestimmt sich der Ort nach den Vorschriften des § 3a Abs. 1 UStG vorbehaltlich der Absätze 2 bis 8 und der §§ 3b und 3e UStG. Wie beim Einzweck- Gutschein unterliegt die Bemessungsgrundlage keiner speziellen Regelung und richtet sich gemäß A 3.17 Abs. 12 UStAE nach den Regelungen des § 10 Abs. 1 S. 2 UStG, folglich nach dem Entgelt. Soweit dem Unternehmer keine Information über die Höhe der erbrachten Gegenleistung vorliegt, greift die Sonderregelung nach § 10 Abs. 1 S. 6 UStG. Dieser Umstand könnte regelmäßig in Vertriebsketten greifen, ist jedoch im Zusammenhang mit der Corona-Pandemie durchaus auch in vielen kleineren Gastronomiebetrieben anzutreffen, wo Gutscheine unterhalb ihres Wertes verkauft wurden, um einen Kaufanreiz und Liquidität zu schaffen. Werden Mehrzweck-Gutscheine nicht eingelöst, ergeben sich hieraus nach A 3.17 Abs. 13 UStAE keine weiteren umsatzsteuerlichen Konsequenzen, da die tatsächliche Leistung nicht erbracht wurde und somit kein steuerbarer Leistungsaustausch stattfand. Die Rückerstattung des Gutscheinwerts an den Kunden löst weiter keine umsatzsteuerlichen Folgen aus, da hier nach A 3.17 Abs. 14 UStAE ein Rücktausch von Zahlungsmitteln erfolgt.

Kritische Analyse

Leistungsort bei Lieferungen in Zusammenhang mit Einzweck-Gutscheinen

[95] Vgl. Heinrichshofen und Kupke, „Gutscheine nach der Änderung der Mehrwertsteuer-Systemrichtlinie (MwStSystRL) - Anmerkungen zu den Auswirkungen auf die deutsche Besteuerungspraxis", S. 144.

[96] Vgl. BFH, „Urt. v. 15.9.2011 - V R 36/09." DStR (2011): 2392-2396, Rn. 17-19.

[97] Vgl. EuGH, „Urt. v. 23.12.2015 - C-250/14, C-289/14 - Air France-KLM, Hop!-Brit Air SAS".

[98] Vgl. Axel Leonard und Martin Robisch, „§ 3 UStG." In Bunjes UStG, Rn. 331; Suse Martin, „§ 3 UStG." In Sölch/Ringleb Umsatzsteuergesetz, 89. EL Juni 2020:Rn. 810.

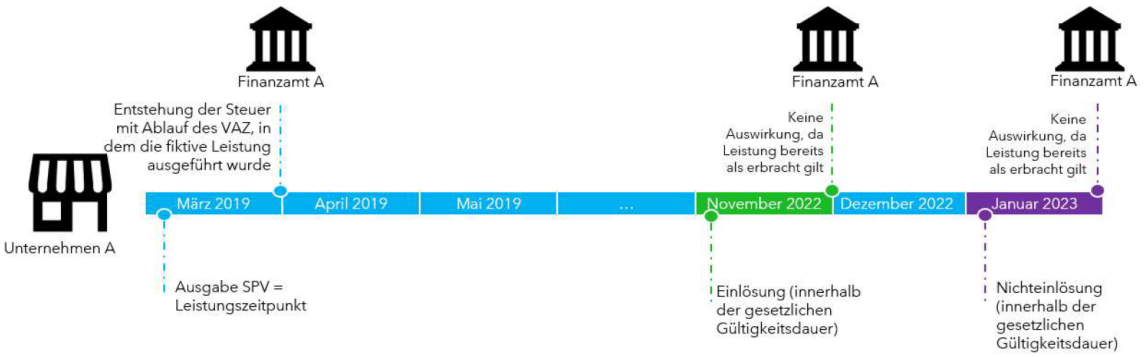

Abbildung 3: Besteuerung von Einzweck-Gutscheinen im eigenen Namen.
Eigene Darstellung.

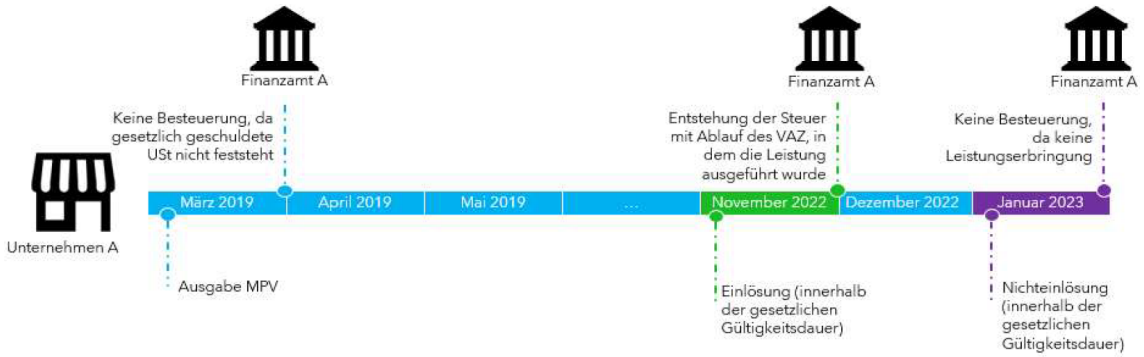

Abbildung 4: Besteuerung von Mehrzweck-Gutscheinen im eigenen Namen.
Eigene Darstellung.

Werden die Regelungen zur Ortsbestimmung näher betrachtet, wird deutlich, dass der Umfang der Fiktion bei der Lieferung und sonstigen Leistung voneinander abweichen. Während die Finanzverwaltung nach A 3.17 Abs. 5 S. 1 bis 3 UStAE bei der Lieferung von einer teilweisen Fiktion ausgeht, ist bei der sonstigen Leistung von einer vollständigen Fiktion auszugehen.[99]

Die Finanzverwaltung nimmt für die fiktive Lieferung eine ruhende Lieferung nach § 3 Abs. 7 UStG an, wodurch eine steuerbefreite Ausfuhr oder innergemeinschaftliche Lieferung bei Ausgabe des Einzweck-Gutscheins nicht zur Anwendung kommt. Dies entspricht nicht dem Wortlaut der zugrundeliegenden Richtlinie. Art. 30b Abs. 1 MwStSystRL besagt, dass „jede Übertragung eines Einzweck-Gutscheins durch einen Steuerpflichtigen, der im eigenen Namen handelt [..] als eine Lieferung [gilt]". Damit werden die allgemein geltenden Regelungen grundsätzlich nicht ausgeschlossen. Folglich wird die Wortwahl der deutschen Finanzverwaltung im Vergleich zum Unionsgesetzgebers kritisch gesehen.[100] Trotzdem begrüßt die herrschende Meinung in der Literatur die Vorgehensweise der Finanzverwaltung aufgrund von Praktikabilitätsgründen und Gründen der Rechtssicherheit.[101] Umsatzsteuerrechtlich ist mit dem Verkauf des Einzweck-Gutscheins die fiktive Lieferung abgeschlossen.

Problematisch kann diese Ansicht der deutschen Finanzverwaltung jedoch im grenzüberschreitenden Lieferverkehr sein, wenn andere Mitgliedstaaten abweichend der nationalen Auffassung von einer vollständigen Fiktion ausgehen.

Beispiel

Ein Kunde kauft bei Unternehmer A in Deutschland unbestritten einen Einzweck-Gutschein in Höhe von 100,00 EUR für 100,00 EUR. Der Zeitpunkt der Besteuerung richtet sich nach § 3 Abs. 14 UStG und der Leistungsort gemäß der teilweisen Fiktion nach A 3.17 Abs. 5 S. 1 UStAE. Der Gutschein wird als Geburtstaggeschenk an einen Freund verschenk. Der beschenkte Freund hat seinen Wohnsitz in Österreich und bestellt Waren bei Unternehmen A. Wird der Liefervorgang ohne Mitwirken des Gutscheins betrachtet, so ist nach § 3c Abs. 1 S. 1 UStG der Leistungsort dort, wo die Versendung endet (= Österreich). Die weiteren Voraussetzungen zur positiven Annahme der Versandhandelsregelung werden als ge-

[99] Vgl. Becker, „Behandlung von Gutscheinen im Umsatzsteuerrecht", S. 226; Denker und Schürkötter, „Umsatzbesteuerung von Gutscheinen in der Praxis", S. 16.

[100] Vgl. Prätzler, „BMF nimmt Stellung zu umsatzsteuerlichen Fragen bei Gutscheinen", S. 2492; Deutscher Industrie- und Handeslkammertag e.V. et al., „Stellungnahme zum Entwurf eines BMF-Schreibens zur umsatzsteuerlichen Behandlung von Einzweck- und Mehrzweck-Gutscheinen", S. 5; Joachim Englisch, „Das neue MwSt-Sonderregime für Gutscheine." ifst-Schrift 515 (2017), S. 76-78.

[101] Vgl. Becker, „Behandlung von Gutscheinen im Umsatzsteuerrecht", S. 226; Denker und Schürkötter, „Umsatzbesteuerung von Gutscheinen in der Praxis", S. 16; Huschens, „Einzweck- und Mehrzweck- Gutscheine - Die zu erwartende Rechtsauffassung der Verwaltung", S. 52.

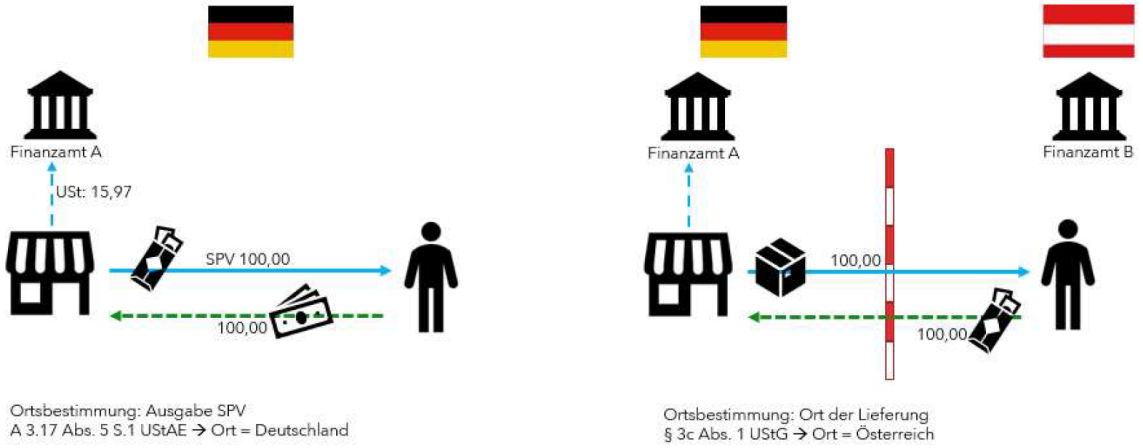

Abbildung 5: Kritische Analyse zur Bestimmung des Leistungsorts bei Lieferungen.
Eigene Darstellung.

geben angesehen. In der Folge müsste Unternehmer A den Umsatz in Österreich versteuern. Aufgrund der teilweisen Fiktion wurde der Umsatz jedoch bereits in Deutschland bei der Ausgabe besteuert. In der Annahme, dass Österreich von einer vollständigen Fiktion ausgeht, wäre ein Besteuerungskonflikt die Folge.

Das dargestellte Beispiel soll die Problematik hinsichtlich des Leistungsorts in vereinfachter Weise darstellen, da die Ausgabe des Einzweck-Gutscheins eigentlich der tatsächlichen Lieferung gleichgestellt werden soll. Problematisch hier ist, dass vorab könnte nicht mehr von einem Einzweck-Gutschein ausgegangen werden, da die Ortsbestimmung regelmäßig von weiteren Handlungen des Leistungsempfängers abhängen würde. Im Ergebnis wäre in der Praxis kaum noch von Einzweck-Gutscheinen auszugehen. Dieser Gedanke entspricht jedoch meiner Ansicht nach nicht der Intention des Gesetzgebers. Der Staat möchte schließlich durch die Leistungsfiktion frühzeitig die Umsatzsteuer vereinnahmen und den Liquiditätsvorteil hier für sich beanspruchen. Aufgrund einer fehlenden Harmonisierung hinsichtlich des Leistungsortes sind daher Doppel- oder Nichtbesteuerung nicht auszuschließen. Eine einheitliche Vorgehensweise ist zwischen den Mitgliedstaaten zu kommunizieren.

Bei einer einheitlichen Regelung besteht für deutsche Versandhändler durch die Fiktion die Chance Steuer zu sparen. Wird Ware an Privatkunden versendet, welche in einem Land mit einem höheren Steuersatz als Deutschland ansässig sind greift grundsätzlich § 3c UStG und der Lieferort ist dort, wo die Versendung endet. Kauft der Kunde jedoch zunächst einen Einzweck-Gutschein, so ist im Zeitpunkt der Ausgabe aufgrund der Fiktion mit dem niedrigeren Steuersatz in Höhe von 19 % zu besteuern. Weiter führt diese Fiktion dann zu Problemen, wenn bei Einlösung eine Zuzahlung zu leisten ist. Die Zuzahlung unterliegt hier dann nicht der Fiktion gemäß dem BMF-Schreiben, sondern müsste unter Umständen als bewegte Lieferung gewürdigt werden.[102] Mit BMF-Schreiben vom 30.6.2020 vertritt die Finanzverwaltung die Ansicht eine Art neue Lieferung in der Zuzahlung zu sehen.[103] Dementsprechend wäre beim Versandhandel ein Teil der Lieferung bereits durch die Leistungsfiktion der Umsatzsteuer unterworfen, wohingegen der übersteigende Teil gegebenenfalls als bewegte Lieferung durch die Versandhandelsregelung im Bestimmungsland zu versteuern ist. Im Ergebnis wird möglicherweise eine Lieferung zwei verschiedenen Steuersätzen unterworfen und der dahinterstehende Umsatz teilweise in Deutschland, teilweise in einem anderen Staat gemeldet. In der Praxis ist hier erhöhe Vorsicht in Zusammenhang mit den Einstellungen in den ERP-Systemen geboten. Zur detaillierten Betrachtung der umsatzsteuerlichen Würdigung von Zuzahlungen im Zusammenhang mit Gutscheinen möchte ich an dieser Stelle auf Kapitel 4.3 verweisen.

Gemäß der Definition von Gutscheinen sind diese nach A 3.17 Abs. 1 S. 2 in unterschiedlicher Form anzutreffen. Das BMF differenziert in der Bestimmung des Leistungsorts nicht nach Gutscheinen in Papierform oder Plastikkarten (physische Gutscheine) und Gutscheinen in elektronischer Form. Eine Vielzahl von Gutscheinen wird jedoch über das Internet verkauft. Diese werden dann in der Regel dem Leistungsempfänger per Mail zugesendet oder können nach Abschluss des Kaufvorgangs heruntergeladen werden. Hier könnte die Frage aufkommen, wo dann eigentlich der Leistungsort ist. Nach § 3 Abs. 7 S. 1 UStG ist dieser dort, wo sich der Gegenstand zur Zeit der Verschaffung der Verfügungsmacht befindet. Die Finanzverwaltung könnte dahingehend verstanden werden, dass der Ort für die Ortsbestimmung maßgeblich ist, wo der Liefergegenstand sich im Zeitpunkt der Ausgabe bzw. Übertragung befindet. Dieser Ort ist jedoch im Zweifelfall nicht definierbar. Eine weitere Interpretationsmöglichkeit ist, dass der Ort der Gutscheinübergabe ausschlaggeben ist.[104] Frag-

[102]Vgl. Daniela Gerhards, „Gutscheine im Umsatzsteuerrecht - BMF-Schreiben vom 2.11.2020." DStZ, Nr. 8 (2021): 307–318, S. 310.

[103]BMF, „Schr. v. 30.6.2020 - III C 2 - S 7030/20/10009 :04, DOK 2020/0610691." BStBl. I (2020): 584, Rn. 30.

[104]Vgl. Becker, „Behandlung von Gutscheinen im Umsatzsteuerrecht", S. 226.

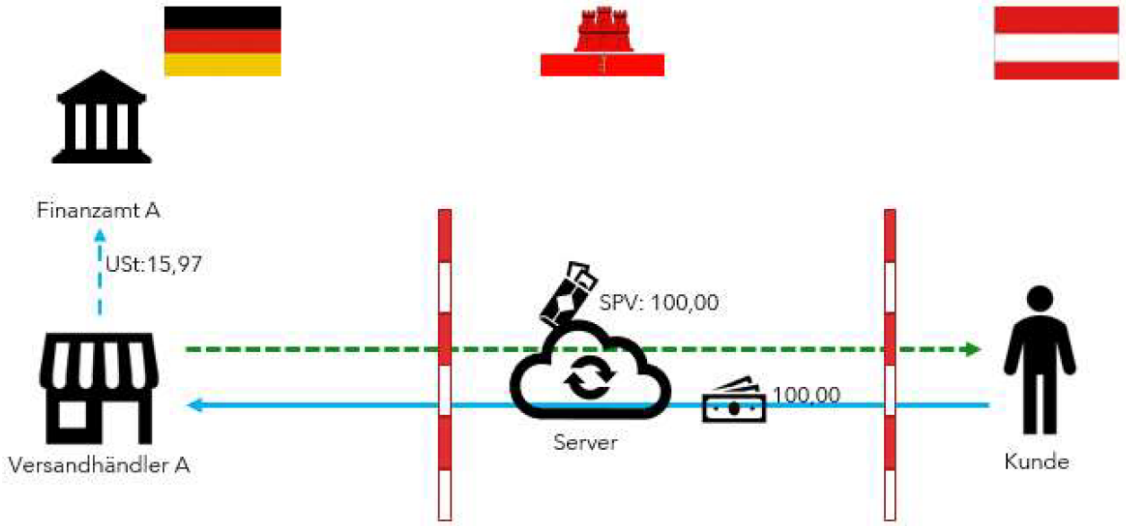

Abbildung 6: Leistungsort bei einem elektronisch generierten Einzweck-Gutschein.
Eigene Darstellung.

lich ist, wo der Ort liegt, wenn der Gutschein als solcher auf einer Homepage gekauft und als Download dem Käufer zur Verfügung gestellt wird.

Beispiel

Ein Versandhändler verkauft an Privatkunden Einzweck-Gutscheine. Diese können im Laden sowie elektronisch über die Homepage erworben werden. Der Versandhändler nutzt hierzu die Serverleistungen von einem Anbieter auf Gibraltar. Mit oben genannten Ausführungen könnte sich der Ort der fiktiven Lieferung nach § 3 Abs. 7 UStG möglicherweise in allen drei Ländern befinden. Deutschland ist der Ort, wo der Liefergegenstand sich im Zeitpunkt der Ausgabe des Gutscheins befindet. Diese Annahme ist dahingehend anzuzweifeln, dass streng genommen eine Vielzahl von potenziellen Gegenständen sich womöglich noch gar nicht beim Versandhändler in Deutschland befinden. Der Ort, wo die Gutscheinausgabe erfolgt, könnte Gibraltar oder Österreich sein. Hier wäre zu klären, ob sich bei einem Download die Gutscheinübergabe dort von statten geht, wo dieser generiert wird oder dieser heruntergeladen wird.

Die Absicht der Finanzverwaltung kann jedoch wohl so interpretiert werden, dass die Besteuerung am Ort des leistenden Unternehmers zu erfolgen hat, dementsprechend ist der Ort der fiktiven Lieferung Deutschland.[105] Eine Konkretisierung seitens der Finanzverwaltung ist bislang nicht erfolgt.

Umkehr der Steuerschuldnerschaft

Das BMF-Schreiben regelt nicht explizit, wie bei Umsätzen zu verfahren ist, die unter das Reverse-Charge-Verfahren gemäß § 13b UStG fallen.[106] Grundsätzlich ist hier nochmal hervorzuheben, dass gemäß Art. 30b MwStSystRL "jede Übertragung eines Einzweck- Gutscheins durch einen Steuerpflichtigen, der im eigenen Namen handelt, [...] als Erbringung der Dienstleistungen, auf die sich der Gutschein bezieht" gilt. Folglich ist bei der Ausgabe eines Gutscheins, welcher sich beispielsweise auf eine sonstige Leistung an einen Unternehmer gemäß § 3a Abs. 2 UStG bezieht, das Reverse-Charge-Verfahren gemäß § 13b Abs. 1 UStG i.V.m. § 13b Abs. 5 S. 1 UStG anzuwenden.

Beispiel

Eine Gärtnerei aus Österreich wird seit Jahren von einem deutschen Unternehmen dazu beauftragt sich um die Pflege der Gartenanlagen des Unternehmens zu kümmern. Es handelt sich unumstritten um eine Dienstleistung (Hecken schneiden, Bäume stutzen, etc.). Der Ort der sonstigen Leistung richtet sich nach § 3a Abs. 2 UStG und ist dort, wo der Empfänger sein Unternehmen betreibt. Unter Vorliegen der sonstigen Voraussetzung des § 1 Abs. 1 Nr. 1 UStG handelt es sich um einen steuerbaren und steuerpflichtigen Umsatz. Die Umkehr der Steuerschuldnerschaft greift nach § 13b Abs. 1 UStG i.V.m. § 13b Abs. 5 S. 1 UStG, da die Gärtnerei eine im übrigen Gemeinschaftsgebiet ansässiger Unternehmer ist. Folglich hat das deutsche Unternehmen die Steuer abzuführen und kann diese gleichzeitig nach § 15 Abs. 1 S. 1 Nr. 4 UStG als Vorsteuer in Abzug bringen. Dementsprechend dürfte es hierzu keine abweichende Anwendung geben, wenn statt der sonstigen Leistung ein Einzweck-Gutschein hingegeben wird. Die Anwendung des Reverse-Charge-Verfahren im Zusammenhang mit der Ausgabe eines Einzweck- Gutscheins ist im genannten Beispiel meiner Meinung nach an sich stimmig und nachvollziehbar. Zweifelsfragen könnten jedoch dann auftreten, wenn die Klassifizierung als Einzweck-Gutschein eindeutig ist, jedoch bei Ausgabe noch nicht feststeht, ob es zum Übergang der Steuerschuld auf den Leis-

[105] Vgl. Becker, „Behandlung von Gutscheinen im Umsatzsteuerrecht", S. 226.

[106] Vgl. Martin Robisch und Stefan Greif, „Entwurf eines "JStG 2018": Die umsatzsteuerlichen Neuerungen aus Sicht der Praxis: Auswirkung und offene Fragen." NWB, Nr. 37 (2018): 2682–2694, S. 2686.

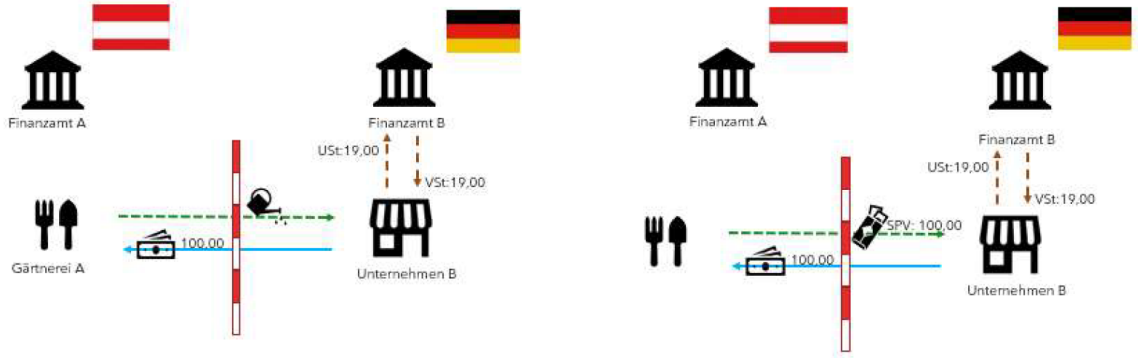

Abbildung 7: Umkehr der Steuerschuldnerschaft bei Einzweck-Gutscheinen.
Eigene Darstellung.

tungsempfänger tatsächlich kommt.

Beispiel

Ein Elektromarkt führt in seinem Sortiment lediglich Waren, die exklusiv dem Regelsteuersatz unterliegen. Ausgegebene Gutscheine sind nur im Inland einlösbar. Wird ein Einzweck-Gutschein an einen Unternehmer ausgegeben, wonach der Gutschein auch für Leistungen, die dem § 13b UStG unterliegen, eingelöst werden, entsteht ein Konflikt. Beispielsweise kann ein Gutschein über 5.000 EUR für Mobiltelefone oder Laptops eingelöst werden. Die Umkehr der Steuerschuld greift nach § 13b Abs. 2 Nr. 10 UStG, soweit die Summe der Entgelte innerhalb eines Vorgangs mindestens 5.000 EUR beträgt. Die Definition eines Einzweck-Gutscheins stellt nicht auf den Steuerschuldner ab, jedoch könnte es hier zu einer unterschiedlichen Behandlung aufgrund der Verwendung des Gutscheins kommen.[107] Eine davon abweichenden Meinung halte ich für bedenkenswert.[108] Insbesondere die Lieferung von Mobilfunkgeräten, Tablet-Computern und Spielekonsolen könnte durchaus auch in Zusammenhang mit Gutscheinen eine Rolle einnehmen. Um diese Konflikte zu vermeiden könnte das leistende Unternehmen dahingehend beraten werden den einzelnen Gutschein lediglich bis zu einer Nennwerthöhe unter 5.000 EUR auszustellen.

Steuerbefreiungen

Die mit dem Leistungsort zusammenhängenden kritischen Punkte können parallel zur von der Finanzverwaltung verweigerten Annahme einer Steuerbefreiung für die Warenbewegung bei der Ausgabe von Einzweck-Gutscheinen angesetzt werden. Die nichtanzuwendende Steuerbefreiung steht zum Widerspruch, dass jeder Umsatz eines Einzweck-Gutscheins wie die eigentliche Leistung zu behandeln ist. Wie auch oben bereits angemerkt, kann es zu einer systemwidrigen Umsatzsteuerbelastung kommen. Würde jedoch eine Steuerbefreiung anwendbar sein, so stellt sich die Frage nach der Nachweispflicht gemäß § 6 Abs. 4 UStG i.V.m. §§ 8 bis 17 UStDV und § 6a Abs. 3 UStG i.V.m. §§ 17a bis 17d UStDV.

Insgesamt äußert sich das BMF hinsichtlich der Steuerbefreiungen lediglich zur Ausfuhrlieferung nach § 4 Nr. 1 Buchst. a UStG sowie der innergemeinschaftlichen Lieferung nach § 4 Nr. 1 Buchst. b UStG. Es ist davon auszugehen, dass alle anderen Steuerbefreiung ohne Anknüpfung an eine Warenbewegung anzuwenden sind.[109] Darunter fallen beispielsweise die Umsätze der Theater, Orchester, botanischen Gärten oder Tierparks nach § 4 Nr. 20 Buchst. a UStG.

Nachträgliche Veränderung der Gegebenheiten

Eine Änderung der Verhältnisse führt regelmäßig zu einer Überprüfung des Status Quo im Hinblick auf die umsatzsteuerliche Würdigung von Lieferungen und sonstigen Leistungen. Im Zusammenhang mit Gutscheinen ist das entscheidendste Kriterium, ob ein Einzweck- oder Mehrzweck-Gutschein vorliegt. Prekär könnte es hinsichtlich der korrekten Besteuerung werden, sofern der Unternehmer sein Sortiment dahingehend ändert, dass die Definition der ausgegebenen Gutscheine sich verändert.[110]

Beispiel

In einem Möbelhaus in Landsberg am Lech erwirbt ein Kunde einen klassischen Nennwert- Gutschein über einen Betrag von 200 EUR für 200 EUR. In diesem Möbelhaus ist im Zeitpunkt der Ausgabe des Gutscheins ausschließlich der Erwerb von Waren, die dem Regelsteuersatz unterliegen, möglich. Da sowohl Ort als auch die Steuerschuld bereits feststehen, liegen die Voraussetzungen für einen Einzweck-Gutschein vor. Die Steuer entsteht nach § 3 Abs. 14 S.1 UStG zum Zeitpunkt der Ausstellung und beträgt 31,93 EUR. Das Möbelhaus entschließt sich nun ausgewählten Teesorten – als

[107] Vgl. Robisch und Greif, „Entwurf eines "JStG 2018": Die umsatzsteuerlichen Neuerungen aus Sicht der Praxis", S. 2686-2687.

[108] Vgl. Denker und Schürkötter, „Umsatzbesteuerung von Gutscheinen in der Praxis", S. 18-19.

[109] Vgl. Robert C. Prätzler, „Umsatzsteuer bei Gutscheinen ab 1.1.2019 - Bestandsaufnahme zu offenen Fragen: Zugleich Anmerkungen zu BMF-Schreiben v. 2.11.2020." StuB, Nr. 3 (2021): 107–112, S. 110.

[110] Vgl. Englisch, „Das neue MwSt-Sonderregime für Gutscheine", S.23; Thiele und König, „Die umsatzsteuerrechtliche Behandlung von Gutscheinen", S. 939.

Ergänzung zu ihrem umfangreichen Sortiment an Teeservice - mit ermäßigtem Steuersatz in das Sortiment aufzunehmen. Nun kann der Kunde zwischen Waren auswählen, die dem regulären als auch ermäßigten Steuersatz unterliegen. Zu diesem Zeitpunkt steht die Höhe der geschuldeten Mehrwertsteuer bei Ausgabe eines Gutscheins nicht weiter fest, wonach nach Definition die jetzt auszugebenden Gutscheine als Mehrzweck-Gutschein zu klassifizieren sind.[111]

Fraglich ist, wie der erstausgebende Gutschein zu werten ist, da dieser theoretisch nun auch für den Erwerb von Tee verwendet werden kann. Denkbar ist ein Wechsel der Klassifizierung vom Einzweck- zum Mehrzweck-Gutschein, wobei die Bemessungsgrundlage und Steuer zu korrigieren sind.[112] Gestützt wird dieser Ansatz durch die umsatzsteuerliche Behandlung von Anzahlungen. Ändert sich hier bei tatsächlicher Ausführung der angezahlten Leistung der Steuersatz, so ist der aktuelle Steuersatz heranzuziehen.[113] Aufgrund der gesetzlichen Fiktion ist dies jedoch bei Einzweck- Gutscheinen nicht möglich.[114] Demgegenüber steht die Alternative, dass es ausschließlich auf die Umstände im Zeitpunkt der Ausgabe des Gutscheins abzustellen sei.[115] Dieser Ansicht ist auch das BMF mit Schreiben vom 30.06.2020 gefolgt, wonach eine spätere Änderung der Verhältnisse unbedeutend ist.[116] Zwar bezieht sich das BMF-Schreiben grundsätzlich auf die temporäre Steuersatzsenkung vom 1. Juli bis 31. Dezember 2020, jedoch dürfte diese Aussage auch auf den oben beschriebenen Sortimentswechsel anzuwenden sein. Gleiches gilt, wenn sich nicht nachträglich die Gutscheindefinition ändert, sondern die Änderung zwischen Ausstellung und Einlösung des Gutscheins auftreten. Das betrifft Kleinunternehmer i.S.d. § 19 UStG, deren Umsätze von der Umsatzbesteuerung ausgeschlossen sind. Dabei könnte jedoch im Zeitpunkt der Gutscheineinlösung der Sonderstatus des Kleinunternehmers nicht mehr vorliegen.

Für die Unternehmer ist bei jeglicher Änderung der Umstände darauf zu achten, dass die betriebsinternen Systeme angepasst und das Personal geschult wird. Führt eine Sortimentserweiterung dazu, dass per Definition statt eines Einzweck-Gutscheins nun ein Mehrzweck-Gutschein vorliegt, ist zu überlegen, wie diese beiden Gutscheinarten visuell bzw. systemseitig in Bezug auf die umsatzsteuerliche Würdigung voneinander abgegrenzt werden können. Dahingehend ist die geforderte Kennzeichnungspflicht eine geeignete Maßnahme. In der Regel besitzen Gutscheine in den Warensystemen auch eigene Warennummern, welche dementsprechend um die „neue" Gutscheinart ergänzt werden können. Weiter muss die dahinterstehende Steuerfindungslogik überprüft und angepasst werden. Der in diesem Zuge verbundene Verwaltungs- und IT-Aufwand darf hierbei nicht unterschätzt werden.

Rechnungsstellung

Werden Gutscheine an einen Unternehmer veräußert, so stellt sich die Frage hinsichtlich der korrekten Rechnungsstellung. Bei der Veräußerung eines Mehrzweck-Gutscheins hat eine Rechnungsstellung aus umsatzsteuerlicher Sicht nicht zu erfolgen, da erst die Einlösung besteuert wird. Sollte hier eine Rechnung mit Ausweis von Umsatzsteuersteuer erfolgen, wird diese nach § 14c Abs. 2 UStG geschuldet. Der Vorsteuerabzug ist nach § 15 Abs. 1 S.1 Nr. 1 UStG ausgeschlossen, da es sich um eine nicht gesetzlich geschuldete Steuer handelt.

Wird jedoch ein Einzweck-Gutschein an einen zum Vorsteuerabzug berechtigten Unternehmer veräußert, stellt sich die Frage nach der korrekten Leistungsbeschreibung nach § 14 Abs. 4 Nr. 5 UStG auf der Rechnung. Das BMF konkretisiert die Anforderung gemäß A 14.5 Abs. 15 S. 5 UStAE dahingehend, dass zum einen die Gutscheinart (Einzweck-Gutschein) und zum anderen eine kurze Beschreibung der Lieferung oder der sonstigen Leistung, zu welcher der Gutschein berechtigt, anzugeben ist. Diese Angabe bereinigt zwar einige für die Praxis relevante Unsicherheiten[117], ist aber dennoch mit Zweifel behaftet. Es kann festgehalten werden, dass ein Einzweck-Gutschein selbst nicht einer ordnungsgemäßen Rechnung nach § 14 Abs. 4 UStG gleichgestellt werden kann.[118] Die Angabe der Gutscheinart ist eine zusätzliche Rechnungsangabe, welche jedoch gesetzlich nicht in §§ 14 Abs. 4, 14a UStG verankert ist.

Einer dementsprechenden Forderung seitens der Bundessteuerberaterkammer, dass die Gutscheinart als Leistungsbeschreibung ausreichend sein sollte, ist daher zuzustimmen. Hier wird folgerichtig auch darauf hingewiesen, dass auch Kleinbetragsrechnungen gem. § 33 S. 1 Nr. 3 UStDV die Angabe über Menge/Art der Leistung erfordern.[119] In aller Regel sind Kleinbetragsrechnungen im Handel oder im Gaststättengewerbe zu finden, welche durch die dort etablierten Kassensysteme generiert werden.[120] Daher sollte die Leistungsbeschreibung dementsprechend kurz ausgestaltet werden können, um diese auch in den Kassensystemen umsetzten zu

[111] In Anlehnung an: Heinrichshofen und Kupke, „Gutscheine nach der Änderung der Mehrwertsteuer- Systemrichtlinie (MwStSystRL) - Anmerkungen zu den Auswirkungen auf die deutsche Besteuerungspraxis", S. 147.

[112] Vgl. Huschens, „Einzweck- und Mehrzweck-Gutscheine in der USt - Praktische Auswirkungen der ab 01.01.2019 geltenden Neuregelungen", S. 452.

[113] Vgl. BMF, „Schr. v. 30.6.2020 - III C 2 - S 7030/20/10009 :04, DOK 2020/0610691", Rn. 12.

[114] Vgl. Rainald Vobbe, „Ausgewählte umsatzsteuerliche Aspekte des "JStG 2018"." BB (2018): 2844–2851, S. 2845.

[115] Vgl. Englisch, „Das neue MwSt-Sonderregime für Gutscheine", S. 23; Bundessteuerberaterkammer, „BMF-Schreiben zur Anwendung der Neuregelungen zu den Gutscheinen", S. 3; Deutscher Industrie- und Handeslkammertag e.V. et al., „Stellungnahme zum Entwurf eines BMF-Schreibens zur umsatzsteuerlichen Behandlung von Einzweck- und Mehrzweck-Gutscheinen", S. 19.2456

[116] Vgl. BMF, „Schr. v. 30.6.2020 - III C 2 - S 7030/20/10009 :04, DOK 2020/0610691", Rn. 30.

[117] Vgl. Becker, „Behandlung von Gutscheinen im Umsatzsteuerrecht", S. 228.

[118] Vgl. Huschens, „Einzweck- und Mehrzweck-Gutscheine - Die zu erwartende Rechtsauffassung der Verwaltung", S. 56.

[119] Vgl. Bundessteuerberaterkammer, „BMF-Schreiben zur Anwendung der Neuregelungen zu den Gutscheinen", S. 4.

[120] Vgl. Jörg Kraeusel, „§ 14 UStG." In Reiß/Kraeusel/Langer UStG (Bonn: Stollfuß), 165. Lieferung:Rn. 265.

können.[121] Mögliche Alternative zur Forderung des BMF ist als Leistungsbeschreibung die Gutscheinnummer anzugeben, um hier eine Verknüpfung zum Gutschein herzustellen. Weiter könnte auch bei einem Einzweck- Gutschein die auf dem Gutschein geforderte Gattung in Kombination mit dem Wort „Gutschein" abgedruckt werden. Dies sollte meiner Ansicht nach als Leistungsbeschreibung ausreichend und nachvollziehbar sein.

Liquidität

Wie oben bereits beschrieben ist das Instrument Gutschein eine nicht unerhebliche Liquiditätsquelle, wobei die Unternehmer insbesondere von der hohen Quote nicht eingelöster Gutscheine profitieren können. Ein weiterer Liquiditätsvorteil wird durch die gesetzliche Neuregelung zur umsatzsteuerlichen Würdigung bei Ausgabe eines Mehrzweck-Gutscheins ermöglicht. Wie dargestellt findet der Leistungsaustausch und somit die Besteuerung erst im Zeitpunkt der Einlösung stattfindet. Bei einer Nichteinlösung ergeben sich, wie in A 3.17 Abs. 13 UStAE festgehalten, keine umsatzsteuerlichen Konsequenzen. Eine Vielzahl von Unternehmen, vornehmlich aus dem Textilbereich, geben aufgrund der gesetzlichen Definition seit 1.1.2019 Einzweck-Gutscheine nach § 3 Abs. 14 UStG aus. Aus betriebswirtschaftlicher Sicht sind daher Gestaltungsmöglichkeiten zu diskutieren, um von der Definition des Einzweck- in den Mehrzweck-Gutschein zu fallen.

Zum einen könnten Unternehmen dahingehend beraten werden ihr Sortiment insoweit zu erweitern, dass die angebotenen Waren oder Dienstleistungen nicht ausschließlich dem regulären oder ermäßigten Steuersatz unterliegen. Die sich daraus ergebende Konsequenz ist, dass nach gesetzlicher Definition kein Einzweck-Gutschein mehr vorliegt. Diese Überlegung könnte jedoch Diskussionen in einer Betriebsprüfung hinsichtlich der Verhältnismäßigkeit aufwerfen.

Beispiel

Ein Schmuckhersteller vertreibt seine Waren sowohl online als auch über eigene Stores und Einzelhändler im Inland. Das Sortiment beinhaltet allein hochwertigen Schmuck sowie Accessoires und Dekorationsartikel, die dem regulären Steuersatz unterliegen. Die ausgegebenen Gutscheine können allein in Deutschland online oder in den eigenen Stores eingelöst werden. Bei Kauf eines Gutscheins wird dieser nicht für ein exklusives Produkt ausgestellt, sondern über einen aufgedruckten Nennwert. Im Gespräch zur Sortimentserweiterung ist ein Buch über das Leben und Wirken der Schmuckfamilie. Bücher unterliegen nach § 12 Abs. 2 Nr. 1 i.V.m. Anlage 2 Lfd. Nr. 49 UStG dem ermäßigten Steuersatz. Es ist davon auszugehen, dass das Buch zum eigentlichen Sortiment nur einen marginalen Umsatzbeitrag leisten wird und Geschenkgutscheine in der Regel als Zahlungsmittel für Waren mit regulärem Steuersatz eingesetzt werden.

Fraglich ist, ob die Finanzverwaltung bei einem solchen Vorgehen einen Missbrauch von rechtlichen Gestaltungsmöglichkeiten nach § 42 AO annehmen könnte. Nichtrechtlich gesprochen ist ein Unternehmer eine „Person, die eine Unternehmung plant, mit Erfolg gründet und/oder selbstständig und verantwortlich mit Initiative leitet, wobei sie persönliches Risiko oder Kapitalrisiko übernimmt".[122] Daher ist grundsätzlich die Gestaltung des Sortiments der Waren und/oder Dienstleistungen allein der Verantwortlichkeit des Unternehmers zuzuordnen. Dementsprechend sind dem Steuerpflichtigen bei der Gestaltung seiner wirtschaftlichen Tätigkeit durch das Gesetz keine Einschränkungen gegeben. Wie bereits angesprochen darf grundsätzlich von Gestaltungsmöglichkeiten Gebrauch gemacht werden. § 42 AO ist dann einschlägig, soweit die Gestaltung unangemessen ist. Unangemessenheit ist dabei grundsätzlich nicht dem Begriff der Ungewöhnlichkeit gleichzusetzten. Die Rechtsprechung hat einige Indizien entwickelt, wonach eine angemessene Gestaltung tendenziell eher einfach, zweckmäßig, übersichtlich und ökonomisch, eine unangemessene Gestaltung hingegen eher unwirtschaftlich, umständlich, kompliziert, schwerfällig, gekünstelt, überflüssig, ineffektiv oder widersinnig erscheint.[123] Auch wenn das angebotene Warensortiment lediglich einen Artikel aufweist, der dem ermäßigten Steuersatz unterliegt, so mag das vielleicht zunächst verwundern, ist jedoch im Wirtschaftsleben auch nicht unüblich. Die sich daraus ergebene Liquiditätsvorteil ist die bloße Rechtsfolge der Gutscheinregelung für Mehrzweck-Gutscheine in § 3 Abs. 15 UStG. Entsprechend bleibt es gemäß EuGH-Rechtsprechung jedem Unternehmer frei, seine wirtschaftliche Tätigkeit in der Art und Weise zu gestalten, dass sich die Umsatzsteuerschuld in Grenzen hält.[124]

Zudem könnte in Betracht gezogen werden die zusätzlich zum Gesetz gestellten Anforderungen seitens der Finanzverwaltung an einen Einzweck-Gutschein nicht anzuwenden. Dazu gehört beispielsweise die Kennzeichnungspflicht oder die Angabe der Gattung. Dieses Vorgehen wäre jedoch mit hohem Risiko behaftet, da grundsätzlich nicht der Umkehrschluss gilt, dass ein nicht entsprechend gekennzeichneter Einzweck- Gutschein kein solcher ist.[125] Die Tatbestandmerkmale ergeben sich schlussendlich aus dem Gesetz und nicht aus er Verwaltungsauffassung.

4.2.2. Umsatzsteuerliche Würdigung innerhalb von Vertriebsketten

Bezüglich der Begriffsdefinition konkretisiert das BMF die Aussteller, Übertragung sowie Ausgabe. In A 3.17 Abs. 1 S. 9 bis 11 UStAE ist geregelt, dass Aussteller derjenige ist, der

[121] Vgl. Bundessteuerberaterkammer, „BMF-Schreiben zur Anwendung der Neuregelungen zu den Gutscheinen", S. 4.

[122] Peter Haric, „Unternehmer: Definition: Was ist "Unternehmer"?". https://wirtschaftslexikon.gabler.de/definition/unternehmer-49340 (letzter Zugriff: 24. März 2021).

[123] Vgl. Eckart Ratschow, „§ 42 AO." In Klein AO, 15. Auflage 2020:Rz. 48.

[124] Vgl. EuGH, „Urt. v. 21.2.2006 - C-255/02 - Halifax." DStR (2006): 420-426, Rz. 73.

[125] Vgl. Prätzler, „Umsatzsteuer bei Gutscheinen ab 1.1.2019 - Bestandsaufnahme zu offenen Fragen", S. 110.

den Gutschein im eigenen Namen ausgestellt hat. Der Verkauf zwischen Unternehmern wird als Übertragung und der der Verkauf eines Gutscheins an einen Kunden als Ausgabe bezeichnet.

Einzweck-Gutscheine in Vertriebsketten

Handeln im eigenen Namen

Das BMF gibt in A 3.17 Abs. 3 UStAE dahingehend Auskunft, wann der Leistungszeitpunkt der fingierten Leistung für Vertriebsketten bei Einzweck-Gutscheinen ist. Hiernach ist der Zeitpunkt der Besteuerung der Leistungsfiktion des leistenden und ausstellenden Unternehmers an den ausgebenden Unternehmer im Zeitpunkt der Übertragung. Die Leistungsfiktion des ausgebenden Unternehmers an den Kunden ist im Zeitpunkt der Ausgabe des Gutscheins an den Kunden. Unabhängig davon ist, ob der Unternehmer als Eigenhändler (in eigenem Namen und für eigene Rechnung) oder als Kommissionär (in eigenem Namen, aber für fremde Rechnung) auftritt.

Ist der ausstellende Unternehmer nicht zugleich auch leistender Unternehmer, so fingiert das Gesetz nach § 3 Abs. 14 S. 3 UStG eine Leistungskette zwischen leistendem Unternehmer, ausstellendem Unternehmer und Kunde. Annähernd an die Grundsätze zur Leistungskommission, gelten beide Leistungen als zeitgleich erbracht.[126] Der Leistungszeitpunkt der fingierten Leistung des leistenden Unternehmers an den Aussteller als auch der fingierten Leistung des ausgebenden Unternehmers an den Erwerber ist nach A 3.17 Abs. 3 S. 5 UStAE der Ausgabezeitpunkt. Damit der leistende Unternehmer eine fristgerechte Besteuerung durchführen kann, ist der Gutscheinaussteller verpflichtet eine Mitteilung hinsichtlich des Zeitpunkts und der Höhe der ausgegebenen Einzweck- Gutscheine an den leistenden Unternehmer zu geben.

Einzweck-Gutscheine in Vertriebsketten (Handeln in fremden Namen)

Wird ein Einzweck-Gutschein in fremden Namen ausgegeben, dann stellt das BMF klar, dass keine Leistungskette fingiert wird. Dies wird gesetzlich bereits durch § 3 Abs. 14 S. 3 UStG geregelt, wonach eine fiktive Leistung desjenigen vorliegt, in dessen Namen die Übertragung des Gutscheins erfolgt. Zeitpunkt der Besteuerung des leistenden Unternehmers ist die Ausgabe des Gutscheins an den Kunden. Damit der leistende Unternehmer seinerseits die dementsprechende Steuerschuld anzeigen und abführen kann, ist wiederrum eine Mitteilung des ausgebenden Unternehmers nach A 3.17 Abs. 4 S. 3 UStAE vorgeschrieben. Die Bemessungsgrundlage richtet sich nach dem vereinnahmten Preis abzüglich der Steuer. Der ausgebende Unternehmer erbringt nach BMF pauschal eine Vermittlungsleistung, bei der die Marge des Unternehmers der Umsatzsteuer unterliegt.

Mehrzweck-Gutscheine in Vertriebsketten

Die Besteuerung von Mehrzweck-Gutscheinen innerhalb von Vertriebsketten ist grundsätzlich aus umsatzsteuerrechtlicher nicht anders zu behandeln wie außerhalb von Vertriebsketten. Auch hier wird lediglich von einem Zahlungsmitteltausch ausgegangen, d.h. dass sowohl die Übertragung von Unternehmen A an Unternehmen B als auch die Ausgabe von Unternehmen B an Kunde C nicht der Umsatzsteuer unterliegt. Handelt der Gutscheinhändler im fremden Namen liegt nach Ansicht der Finanzverwaltung eine Vergütung für eine steuerbare Vermittlungsleistung des Gutscheinhändlers an den leistenden Unternehmer vor. Handelt Unternehmen B im Namen des ausstellenden bzw. übertragenden Unternehmers A, so ist nach A 3.17 Abs. 11 S. 5 UStAE im Zeitpunkt der Übertragung und Ausgabe an den Kunden C eine Vermittlungsleistung erbracht worden. Der Ort der Vermittlungsleistung bestimmt sich nach § 3a Abs. 2 UStG. Sofern der Gutscheinhändler in eigenen Namen handelt, ist grundsätzlich von einer steuerbaren Leistung an den ausstellenden Unternehmer auszugehen.

Kritische Analyse

Begrifflichkeiten

Übergreifend lässt sich hinsichtlich der Begriffsdefinitionen eine Abweichung zum Gesetzestext erkennen. Dort ist in § 3 Abs. 14 S.2 und 3 UStG festgehalten, dass die Übertragung eines Gutscheins einer Leistungsfiktion unterliegt. Diese Fiktion bezieht sich auf den Verkauf an den Endkunden. Dem steht die Definition des BMF gegenüber, dass eine Übertragung den Verkauf eines Gutscheins zwischen Unternehmern darstellt. Das hätte zur Folge, dass der Verkauf, sprich die Ausgabe eines Gutscheins, an einen Kunden gar nicht der Leistungsfiktion unterliegt. Der Ansicht, dass die Begriffsdefinition seitens des BMF keine Einschränkung der gesetzlichen Norm darstellen soll, ist zuzustimmen.[127] Jedoch ist eine Anpassung der Wortwahl im BMF-Schreiben naheliegend, da hier eine widersprüchliche Rechtsauffassung vorliegt.

Nachträgliche Veränderung der Gegebenheiten

Wie bereits unter Punkt 1.3.4 angesprochen, stellt sich auch bei der Ausgabe von Gutscheinen innerhalb von Vertriebsketten die Frage, wie mit Gutscheinen zu verfahren ist, die beispielsweise vor einem Sortimentswechsel ausgegeben und dementsprechend nicht mehr als Einzweck-Gutschein

[126] Vgl. Robert Hammerl und Andreas Fietz, „Umsatzsteuerliche Behandlung von Einzweck- und Mehrzweck- Gutscheinen: BMF klärt mit Schreiben v. 2.11.2020 Anwendungsfragen." NWB, Nr. 48: 3538–3549, S. 3544.

[127] Vgl. Becker, „Behandlung von Gutscheinen im Umsatzsteuerrecht", S. 225.

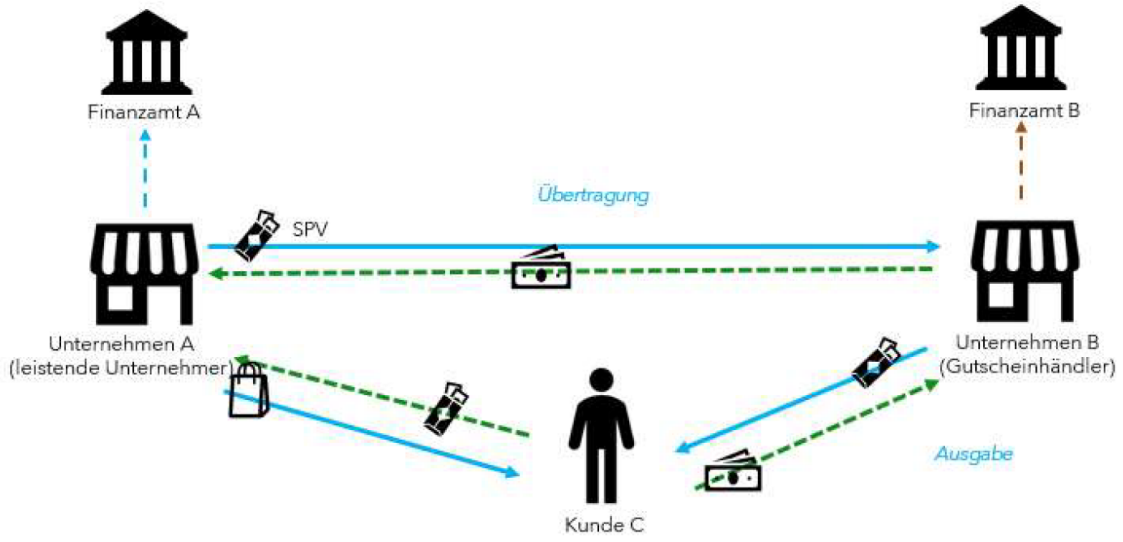

Abbildung 8: Einzweck-Gutscheine in Vertriebsketten (Handeln im eigenen Namen).
Eigene Darstellung.

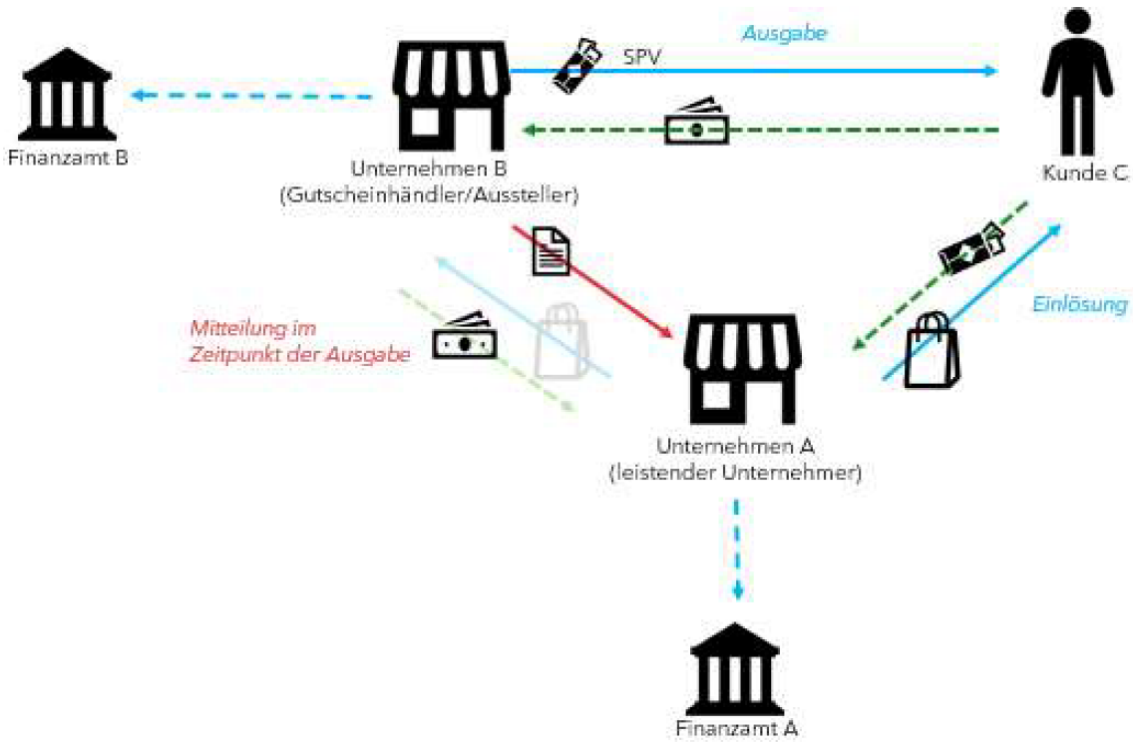

Abbildung 9: Einzweck-Gutscheine in Vertriebsketten (Handeln in eigenen Namen) bei abweichendem leistendem Unternehmer.
Eigene Darstellung.

zu klassifizieren sind. Da das BMF-Schreiben hierzu den maßgeblichen Leistungszeitpunkt die Ausgabe vorgibt, sollte m.E. durch eine nachträgliche Änderung keine Anpassung hinsichtlich der Besteuerung mehr vorzunehmen sein.

Meldepflicht bei Einzweck-Gutscheinen in Vertriebsketten

Damit die Meldepflicht des leistenden Unternehmers der fingierten Leistung im Ausgabezeitpunkt ordnungsgemäß erfolgen kann, fordert die Finanzverwaltung eine dementsprechende Mitteilung des ausstellenden Unternehmers an den leistenden Unternehmer. Das bedeutet, dass der leistende Unternehmer hinsichtlich seiner Steuerschuld vom ausstellenden Unternehmer abhängig ist. Fraglich ist, welche Rechtsfol-

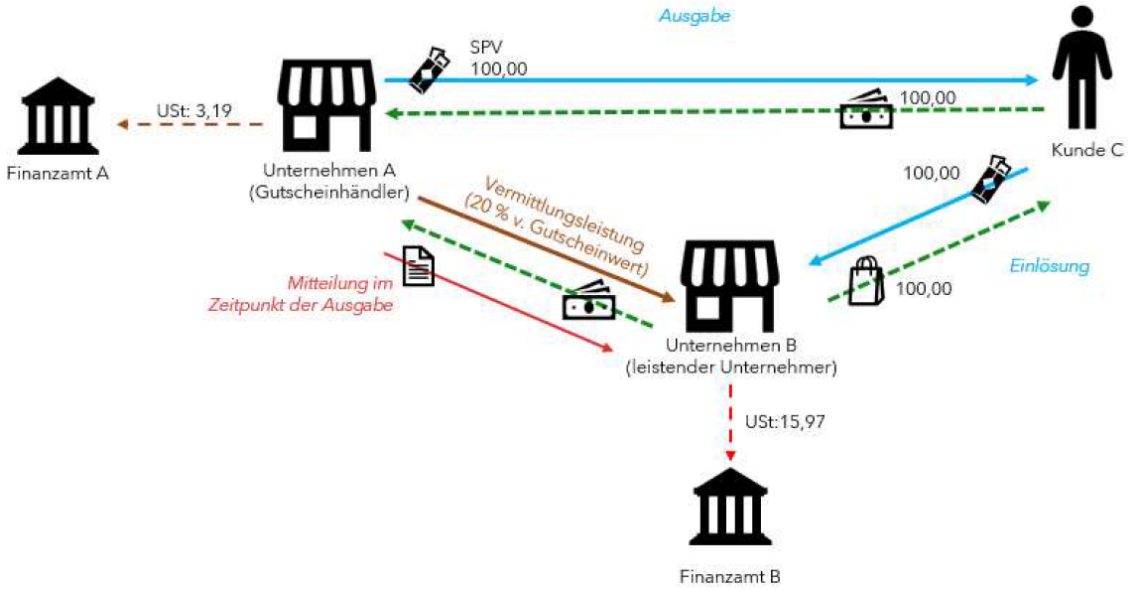

Abbildung 10: Einzweck-Gutscheine in Vertriebsketten (Handeln in fremden Namen).
Eigene Darstellung.

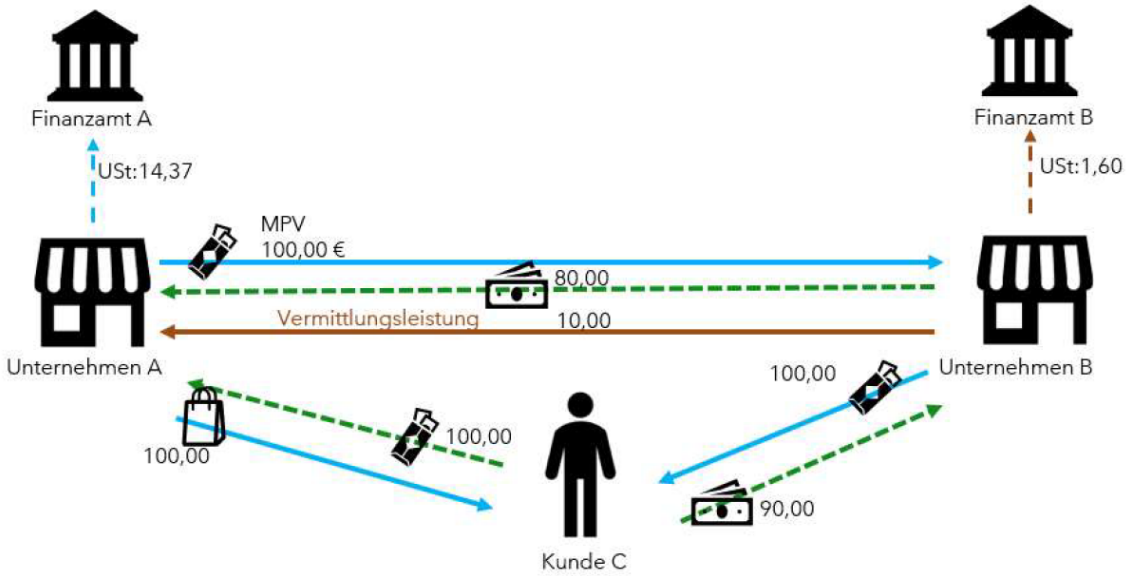

Abbildung 11: Mehrzweck-Gutscheine in Vertriebsketten.
Eigene Darstellung.

gen eintreten, wenn diese Mitteilung unvollständig ist oder insgesamt unterlassen wird. Im Zweifelsfall erfährt der leistende Unternehmer erst von seiner Steuerschuld, wenn der jeweilige Gutschein bei ihm eingelöst wird. Jedoch entsteht die Steuerschuld unabhängig von der Kenntnislage im Ausgabezeitpunkt an den Kunden. Folglich müssen solche Pflichten vertraglich geregelt werden, um daraus entsprechende finanzielle Nachteile zivilrechtlich gegen den Gutscheinaussteller zu begründen.[128] Die Form der Mitteilung wird seitens der Finanzverwaltung nicht spezifiziert. Daher könnte bei Anwendung des umsatzsteuerlichen Gutschriftverfahrens gemäß § 14 Abs. 2 S. 2 UStG die Rechnung im Regelfall die Funktion als Mitteilung innehaben.[129]

Annahme einer Vermittlungsleistung bei Mehrzweck-Gutscheinen

Das BMF sieht pauschal eine Vermittlungsleistung bei Vertriebsketten mit Mehrzweck- Gutscheinen. Diese Auffassung

[128]Vgl. Huschens, „Einzweck- und Mehrzweck-Gutscheine - Die zu erwartende Rechtsauffassung der Verwaltung", S. 51.

[129]Vgl. Hammerl und Fietz, „Umsatzsteuerliche Behandlung von Einzweck- und Mehrzweck-Gutscheinen", S. 3544.

wird jedoch ungeachtet der Frage nach den zugrunde liegenden vertraglichen Vereinbarungen getroffen.

Insbesondere die Annahme einer Vermittlungsleistung bei einem Mehrzweck-Gutschein im eigenen Namen scheint fraglich zu sein. Das BMF spricht zwar in A 3.17 Abs. 10 S. 3 UStAE nur noch von einer „steuerbare[n] Leistung", jedoch ist nicht erkennbar, worin diese konkret liegen soll, wobei im Entwurf noch das Wort „Vermittlungsleistung" genannt wurde. Art. 30b Abs. 2 MwStSystRL spricht hier von Vertriebs- oder Absatzförderungsleistungen. Vergleicht man die Ausführungen zu Vermittlungsleistungen in A 3.7 Abs. 1 S. 3 UStAE ist eine verallgemeinerte Annahme seitens der Finanzverwaltung inkonsistent. Eine entsprechende Begründung ist im BMF-Schreiben[130] nicht genannt. Insbesondere in Fällen, in denen ein Unternehmer Mehrzweck-Gutscheine erwirbt, um sie auf eigene Rechnung weiter an Kunden zu veräußern, liegt m.E. und nach Literaturmeinung keine Vermittlungsleistung vor.[131] Zur Schaffung der Rechtssicherheit sollte hier eine Klarstellung dahingehend erfolgen, dass nicht pauschal von einer Vermittlungsleistung auszugehen ist.[132] Diese von der Finanzverwaltung abweichende Beurteilung könnte im Falle einer Betriebsprüfung vom Zwischenhändler durch Vertragsunterlagen oder dem dahinterstehenden Geschäftsvorgang dargestellt werden können.

Für die Vermittlungsleistung müsste sich eine Pflicht zur Erteilung einer Rechnung nach § 14 Abs. 2 UStG ergeben, wobei davon auszugehen ist, dass diese lediglich einen informellen Charakter aufweist. Dabei erstaunt es, dass das BMF hinsichtlich der Vermittlungsleistung nicht auf die Pflicht zur Rechnungsstellung nicht eingeht.[133]

Grenzüberschreitende Vertriebsketten bei Mehrzweck-Gutscheinen

Der Ort der Vermittlungsleistung ist nach Auffassung der Finanzverwaltung nach § 3a Abs. 2 UStG zu bestimmen. Das heißt der Ort ist am Sitz des Leistungsempfängers. In dem Fall, dass Gutscheinaussteller und Gutscheinhändler ihren Sitz in unterschiedlichen Mitgliedsstaaten haben, müsst es regelmäßig zu einer Umkehr der Steuerschuld nach § 13b Abs. 1 i.V.m § 13b Abs. 5 UStG kommen. Problematisch könnte es in der Praxis dahingehend werden, dass der Gutscheinaussteller zum Steuerschuldner für eine Leistung wird, die im womöglich nicht umfassend bekannt ist und im Zweifel der Höhe nach nicht einmal zu definieren ist. Ohne eine Information seitens des Gutscheinhändlers könnte eine Besteuerung dieser Vermittlung nicht erfolgen.[134]

4.3. Temporäre Senkung der Steuersätze

Die mit dem Zweiten Corona-Steuerhilfegesetz[135] beschlossene temporäre Steuersatzsenkung für den Zeitraum vom 1.7.2020 bis 31.12.2020 auf 16 % bzw. 5 % wirkt sich demzufolge auch auf die umsatzsteuerliche Behandlung von Einzweck-Gutscheinen aus. Die Ausgabe und das Einlösen von Mehrzweck-Gutscheinen sind aufgrund der oben beschriebenen Behandlung unkritisch. Übergreifendes Ziel der Steuersatzsenkung ist die Kaufkraft zu stärken und dementsprechend die Wirtschaft zu beflügeln.[136] Ein Anreiz sollte dahingehend geschaffen werden, Geld nicht etwa zu sparen, sondern in hochwertige Konsumgüter oder Dienstleistungen zu investieren.[137] Jeweils am 30.6.2020[138] und 4.11.2020[139] wurden BMF-Schreiben zur Steuersatzsenkung veröffentlicht, die unter anderem Stellung hinsichtlich der Auswirkungen auf die Einlösung von Einzweck- Gutscheinen nehmen. Trotzdem bleiben die in diesem Zeitraum getätigte Umsätze mit Einzweck-Gutscheinen problembehaftet.

4.3.1. Abgrenzung einer Anzahlung von einem Gutschein

Die Chance mit der Ausgabe von Einzweck-Gutscheinen die temporäre Steuersatzsenkung über das Jahresende zu sichern, wird in einem Beitrag von Kathrin Feil und Ursula Slapio diskutiert. Dabei wird verkürzt das folgende Szenario dargestellt: Die beschlossene Steuersatzsenkung führt bei einem Unternehmer zu einer massiven Nachfragesteigerung, die bis zum Jahresende tatsächlich nicht mehr bedient werden kann.[140] Infolgedessen unterliegen diese Leistung nach § 13 Abs. 1 Nr. 1 Buchst. a UStG wieder dem erhöhten Steuersatz, da die Entstehung der Steuer an den Zeitpunkt der Leistungserbringung (Verwirklichung des Vorgangs) anknüpft.[141] Auch eine vorzeitige Zahlung bis Ablauf des Jahres 2020 würde kein anderes Ergebnis erzielen. Zwar entsteht die Steuer gemäß § 13 Abs. 1 Nr. 1 Buchst. a S. 4 UStG im Zeitpunkt der Vereinnahmung, jedoch müsste diese nach § 27 Abs. 1 S. 2,3

[130] Vgl. BMF, „Schr. v. 2.11.2020 - III C 2 - S 7100/19/10001 :002, DOK 2020/1051603", S. 9.

[131] Vgl. André Meißner, „Praxisprobleme bei Mehrzweckgutscheinen in (grenzüberschreitenden) Vertriebsketten: Zugleich eine erste Auseinandersetzung mit dem BMF-Schreiben v. 2.11.2020." UR (2020): 914–918, S. 916; Deutscher Industrie- und Handeslkammertag e.V. et al., „Stellungnahme zum Entwurf eines BMF-Schreibens zur umsatzsteuerlichen Behandlung von Einzweck- und Mehrzweck- Gutscheinen", S. 11.

[132] Vgl. Deutscher Industrie- und Handeslkammertag e.V. et al., „Stellungnahme zum Entwurf eines BMF- Schreibens zur umsatzsteuerlichen Behandlung von Einzweck- und Mehrzweck-Gutscheinen", S. 11-13.

[133] Vgl. Gerhards, „Gutscheine im Umsatzsteuerrecht - BMF-Schreiben vom 2.11.2020", S. 311.

[134] Vgl. Meißner, „Praxisprobleme bei Mehrzweckgutscheinen in (grenzüberschreitenden) Vertriebsketten", S. 917.

[135] Deutscher Bundestag, „Zweites Gesetz zur Umsetzung steuerlicher Hilfemaßnahmen zur Bewältigung der Corona-Krise (Zweites Corona-Steuerhilfegesetz)." BGBl. I Nr. 31 (2020): 1512–1516.

[136] Vgl. Deutscher Bundestag, „Gesetzentwurf der Fraktionen der CDU/CSU und SPD: Entwurf eines Zweiten Gesetzes zur Umsetzung steuerlicher Hilfsmaßnahmen zur Bewältigung der Corona-Krise (Zweites Corona- Steuerhilfegesetz)." BT-Drs. 19/20058, S. 1.

[137] Vgl. Andreas Erdbrügger, „Die temporäre Senkung der Umsatzsteuersätze - Fluch oder Segen?" UR (2020): 485–491, S. 486.

[138] BMF, „Schr. v. 30.6.2020 - III C 2 - S 7030/20/10009 :04, DOK 2020/0610691", Rn. 30.

[139] BMF, „Schr. v. 4.11.2020 - III C 2- S 7030/20/10009 :016, DOK 2020/1074476." BStBl. I (2020): 1129, Rn. 5-6.

[140] Vgl. Ursula Slapio und Kathrin Feil, „Temporäre Senkung der Mehrwertsteuersätze - Gestaltungsmöglichkeiten durch Ausgabe von Einzweck-Gutscheinen?" UR (2020): 703–706, S. 703-704.

[141] Vgl. Carsten Leipold, „§ 13 UStG." In Sölch/Ringleb Umsatzsteuergesetz, Rn. 40.

UStG auf den Steuersatz der tatsächlichen Leistungserbringung korrigiert werden.

Daher wird in Erwägung gezogen mit Hilfe der gesetzlichen Regelung zur Besteuerung von Einzweck-Gutscheinen den temporär gesenkten Steuersatz für eine bereits bestellte bzw. in Auftrag gegebene Leistung zu sichern, auch wenn die tatsächliche Leistungserbringung erst nach dem 31.12.2020 erfolgt. Unter der Annahme, dass die Kriterien eines Einzweck- Gutscheins nach § 3 Abs. 14 UStG vorliegen, wird dieser bei Erwerb als fiktive Leistung behandelt und wäre dementsprechend dem ermäßigten Steuersatz zu unterwerfen. Diese Möglichkeit wäre konform mit der Systematik des Umsatzsteuergesetzes und würde sich m.E. auch in der Zielsetzung des Gesetzgebers wiederfinden.[142] Wie angedeutet ist die Intension, dass Konsumenten in hochpreisige Leistungen investieren, eine steigernde Auswirkung auf Konsumgüter und Dienstleistungen des täglichen Bedarfs dürfte kaum von Relevanz sein.[143] Wirtschaftlich gesehen sollte die gesenkte Umsatzsteuer folglich auch den Konsumenten zugutekommen, die sich grundsätzlich innerhalb dieses Zeitfensters für eine Investition entschieden haben.

Das BMF vertritt dagegen eine konträre Auffassung. Im Schreiben vom 4.11.2020 vertritt die Finanzverwaltung die Auffassung, dass die „Ausstellung eines als Gutschein bezeichneten Dokuments für einen verbindlich bestellten Gegenstand, bei dem ein späterer Umtausch, eine Barauszahlung oder eine Übertragung des Gutscheins auf einen anderen Verkäufer bzw. Käufer ausgeschlossen ist und dessen Ausstellung mit einer Abnahmeverpflichtung verbunden ist, ist eine Anzahlung." Weiter differenziert die Finanzverwaltung einen Einzweck-Gutschein und eine Anzahlung bezugnehmend auf den Leistungsgegenstand. Dabei sei der Leistungsgegenstand beim Einzweck-Gutschein lediglich bestimmbar, wohingegen bei der Anzahlung dieser inhaltlich genau bestimmt ist. Zudem hätte der Erwerber keinen Einfluss den Leistungszeitpunkt zu wählen oder zu beeinflussen.[144] Das als „Gutschein" benannte Dokument für einen verbindlich bestellten Gegenstand sei als eine Art Zahlungsnachweis anzusehen.[145]

Fraglich ist, ob die Rechtsauffassung der Finanzverwaltung richtig ist. Werden die jeweiligen Voraussetzungen miteinander verglichen, so fordern beide, dass der Ort der Leistung und die geschuldete Steuer feststehen. Zudem hat die zu erbringende Leistung zumindest bestimmbar zu sein. Die vom BMF dargelegten Kriterien finden sich nicht im Gesetz wieder. Die Ausstellung eines Gutscheins mit einer konkreten Leistung dürfte in der Praxis nicht unüblich sein. Weswegen dies im Zusammenhang mit der Steuersatzsenkung und eines verbindlich bestellten Neuwagens oder einer Werklieferung ausgeschlossen ist, ist nicht nachvollziehbar. Aufgrund der dargelegten Ähnlichkeiten hinsichtlich der Kriterien ist davon auszugehen, dass sich die Finanzgerichte und der BFH mit der Abgrenzung möglicherweise näher befassen müssen, da die Meinung in der Literatur vertreten wird, dass sich die Finanzämter kreativ zeigen könnten, um die Besteuerung mit 16 % zu verneinen.[146]

4.3.2. Restaurantdienstleistungen

Eine Unterscheidung zwischen dem Regelsteuersatz unterliegenden Restaurations- oder Verpflegungsdienstleistung und den ermäßigt besteuerten Lieferungen von Nahrungsmitteln ist oftmals problembehaftet.[147] Eine Restaurant- und Verpflegungsdienstleistung wird in Art. 6 Abs 1 MwStVO als Abgabe zubereiteter oder nicht zubereiteter Spiesen oder Getränke in Kombination mit ausreichenden unterstützenden Dienstleistungen, wobei dieser in der Gesamtbetrachtung überwiegt, definiert. Der Großteil gastronomischer Betriebe gibt zudem Restaurantgutscheine aus. Diese wurden im ersten Lockdown im Frühjahr 2020 sowie im zweiten Lockdown ab November 2020 von den Gastwirten intensiv beworben, um Liquiditätseinbußen aufgrund eingebrochener Umsätze teilweise kompensieren zu können.

Für die durch die Corona-Krise stark betroffene Gastronomiebranche wurde bereits mit dem mit ersten Corona-Steuerhilfegesetz eine Umsatzsteuerermäßigung von 19 % auf 7 % vom 1.7.2020 bis 30.6.2021 für die Abgabe von Speisen beschlossen.[148] Weiter wurden diese noch aufgrund der branchenübergreifenden Steuersatzsenkung angepasst.[149] Mit dem dritten Corona-Steuerhilfegesetz wurde die Anwendung des ermäßigten Steuersatzes auf die Speiseabgabe bis 31.12.2022 verlängert.[150] Die folgende Tabelle 1 soll eine Übersicht geben, wann für welchen Umsatz welcher Steuersatz anzuwenden ist.

Dabei war es der Finanzverwaltung offensichtlich ein Anliegen explizit die Würdigung von Gutscheinen für Restaurationsleistungen im Schreiben vom 4.11.2020[151] aufzuführen. Im Zeitpunkt der Veröffentlichung dieses Schreibens war eine weitere Fristverlängerung noch nicht beschlossen. Daher wird im Folgenden die im BMF-Schreiben genannte Frist bis

[142] Vgl. Michael Rust und Jochen Tillmanns, „Zu der umsatzsteuerlichen Behandlung von Einzweck- Gutscheinen im Lichte der vorübergehenden Mehrwertsteuersenkung." USt direkt digital, Nr. 21 (2020): 17–19, S. 18.

[143] Erdbrügger, „Die temporäre Senkung der Umsatzsteuersätze - Fluch oder Segen?", S. 486.

[144] Vgl. BMF, „Schr. v. 4.11.2020 - III C 2- S 7030/20/10009 :016, DOK 2020/1074476", Rn. 5.

[145] Vgl. Rust und Tillmanns, „Zu der umsatzsteuerlichen Behandlung von Einzweck-Gutscheinen im Lichte der vorübergehenden Mehrwertsteuersenkung", S. 17.

[146] Vgl. Rust und Tillmanns, „Zu der umsatzsteuerlichen Behandlung von Einzweck-Gutscheinen im Lichte der vorübergehenden Mehrwertsteuersenkung", S. 19.

[147] Vgl. Christoph Wäger, „§ 3a UStG." In Sölch/Ringleb Umsatzsteuergesetz (München: C.H. Beck oHG), 90. EL:Rn. 174.

[148] Vgl. Deutscher Bundestag, „Gesetz zur Umsetzung steuerlicher Hilfsmaßnahmen zur Bewältigung der Corona-Krise (Corona-Steuerhilfegesetz)." BGBl. I Nr. 30: 1385-1386, S. 1385.

[149] Deutscher Bundestag, „Zweites Gesetz zur Umsetzung steuerlicher Hilfemaßnahmen zur Bewältigung der Corona-Krise (Zweites Corona-Steuerhilfegesetz)", S. 1514.

[150] Vgl. Deutscher Bundestag, „Drittes Gesetz zur Umsetzung steuerlicher Hilfsmaßnahmen zur Bewältigung der Corona-Krise." BGBl. I Nr. 10 (2021): 330-331.

[151] Vgl. BMF, „Schr. v. 4.11.2020 - III C 2- S 7030/20/10009 :016, DOK 2020/1074476", Rn. 6.

Tabelle 1: Übersicht der geltenden Steuersätze bei Restaurantdienstleistungen.
Vgl. Rüdiger Weimann, „Umsatzsteuer: Befristete Absenkung der USt für Gastronomen - Checkliste, Fälle und Praxistipps." BBP (2020): 179–183, S. 179.

	bis 30.06.20	1.7.20 – 31.12.20	1.1.21 – 31.12.22	ab 01.01.23
Speisen zum Verzehr an Ort und Stelle	19%	5%	7%	19%
Speisen „to go"	7%	5%	7%	7%
Getränke (Grundsatz)	19%	16%	19%	19%
Getränke (Milch und Milchmischgetränke)	7%	5%	7%	7%

1.7.2021 verwendet. Eine dahingehende Korrektur ist zum heutigen Zeitpunkt noch nicht erfolgt, jedoch sollte die Angabe 1.7.2021 problemlos auf 1.1.2023 angepasst werden können.

Demnach sollen Gutscheine, welche vor dem 1.7.2020 ausgegeben wurden als Einzweck- Gutscheine gewürdigt werden, auch wenn deren Einlösung im Zeitraum vom 1.7.2020 bis 30.6.2021 erfolgt. Aufgrund der unterschiedlichen Steuersätze für Speisen und Getränke geht die Finanzverwaltung bei der Ausgabe von Gutscheinen grundsätzlich von Mehrzweck-Gutscheinen aus, mit der Folge, dass diese im Zeitpunkt der tatsächlichen Leistungserbringung und dem dann gültigen Steuersatz versteuert werden. Eine Klassifizierung als Einzweck-Gutschein im Zeitraum vom 1.7.2020 bis 30.6.2021 sei nur dann möglich, wenn der Gutschein sich explizit auf den Bezug von Speisen oder Getränken bezieht.[152]

Zusammenfassend hängt laut Finanzverwaltung die umsatzsteuerliche Würdigung der Ausstellung und Einlösung des Gutscheins davon ab, ob

- der Gutschein sowohl für Speisen als auch für Getränke gilt
- der Gutschein ausschließlich für Speisen bzw. ausschließlich für Getränke gilt
- der Gutschein unbefristet gilt oder bis zu einem bestimmten Zeitpunkt eingelöst werden muss.

Regelmäßig werden im Gastronomiebereich zwei Arten von Gutscheinen ausgegeben. Zum einen die klassischen Nennwertwertgutscheine über einen bestimmten Betrag, welcher in der Regel für Speisen als auch für Getränke gilt. Zum anderen werden oftmals Gutscheine über Aktionen wie z.B. Themenbuffets oder Frühstücksbrunch ausgestellt. Bei zweiteren sind die beinhalteten Leistungen und deren Steuer insoweit dem Grunde nach bestimmt, sodass § 3 Abs. 14 UStG greift und ein Einzweck-Gutschein vorliegt. Der Ort der Restaurationsleistung bestimmt sich nach § 3a Abs. 3 Nr. 3 Buchst. b UStG und liegt dort, wo sie tatsächlich vom Unternehmer erbracht wird.

Folglich ist bei einer Gestaltung dahingehend, dass Nennwertgutscheine nur für Speisen ausgegeben werden, ein Umsatzsteuer-Vorteil von bis zu 14 % möglich. Insbesondere gegen Ende der Frist, könnte dieser Vorteil genutzt werden. Jedoch werden die meisten Gastronomen kaum einen Nennwertgutschein ausgeben und diesen auf das Essen vor Ort beschränken. In der Corona-Krise wurde das Essen-to-go ein Instrument, um einen Bruchteil der normalen Umsätze generieren zu können und wurde von den Kunden auch angenommen. Daher dürften in der Praxis dem Grunde nach regelmäßig Mehrzweck- Gutscheine vorliegen.

4.3.3. Zuzahlungen bei Einlösung des Einzweck-Gutscheins

Weitere Unklarheit hinsichtlich Einzweck-Gutscheine bestehen, wenn seitens des Kunden eine Zuzahlung bei Einlösung nach dem 31.12.2020 geleistet wird. Zweifelhaft ist, ob diese Zuzahlung dann mit dem aktuell gültigen Steuersatz oder dem zum Zeitpunkt des Kaufs des Gutscheins gültigen Steuersatz zu besteuern ist. Nach Ansicht der Finanzverwaltung ist die Zuzahlung des Gutscheininhabers mit dem im Zeitpunkt der Gutscheineinlösung geltenden Steuersatz zu versteuern. Diese Auffassung wird jedoch ohne weiterführende Begründung dargelegt.[153] M.E. stößt diese Ansicht zu Recht auf kritische Stimmen in der Literatur.[154] Wie bereits öfters erwähnt wird in § 3 Abs. 14 S. 2 UStG eine Fiktion erschaffen, dass die mit dem Gutschein zusammenstehende Leistung als erbracht gilt. Folgt man dementsprechend der Auffassung des BMF, so müsste davon auszugehen sein, dass die Zuzahlung mit einem anderen Steuersatz als eigenständige Leistung i.S.d. § 3 UStG zu würdigen ist. Jedoch wird kein weiterer Gegenstand geliefert.[155] Die Zuzahlung müsste eigentlich eine Änderung der Bemessungsgrundlage nach § 17 Abs. 1 S. 1 UStG darstellen, da diese sowohl eine nachträgliche Entgeltminderung als auch -erhöhung umfasst.[156]

Richtig ist jedoch, dass eine Behandlung nach § 17 UStG weitreichende Gestaltungsmöglichkeiten ermöglicht.[157] Wird ein Einzweck-Gutschein zum gesenkten Steuersatz gekauft

[152] Vgl. BMF, „Schr. v. 4.11.2020 - III C 2- S 7030/20/10009 :016, DOK 2020/1074476", Rn. 6.

[153] Vgl. BMF, „Schr. v. 30.6.2020 - III C 2 - S 7030/20/10009 :04, DOK 2020/0610691", Rn. 30.

[154] Vgl. Karl-Hermann Eckert, „Die Gutscheinregelung im Zusammenhang mit der Absenkung der Umsatzsteuer." USt direkt digital, Nr. 19 (2020): 15–16, S. 16; Rust und Tillmanns, „Zu der umsatzsteuerlichen Behandlung von Einzweck-Gutscheinen im Lichte der vorübergehenden Mehrwertsteuersenkung", S. 19; Slapio und Feil, „Temporäre Senkung der Mehrwertsteuersätze - Gestaltungsmöglichkeiten durch Ausgabe von Einzweck-Gutscheinen?", S. 704; Gerhards, „Gutscheine im Umsatzsteuerrecht - BMF-Schreiben vom 2.11.2020", S. 317.

[155] Slapio und Feil, „Temporäre Senkung der Mehrwertsteuersätze - Gestaltungsmöglichkeiten durch Ausgabe von Einzweck-Gutscheinen?", S. 704.

[156] Vgl. Christian Korn, „§ 17 UStG." In Bunjes UStG (München: C.H. Beck oHG), Rn. 24.

[157] Vgl. Rust und Tillmanns, „Zu der umsatzsteuerlichen Behandlung von Einzweck-Gutscheinen im Lichte der vorübergehenden Mehrwertsteuersenkung", S. 19.

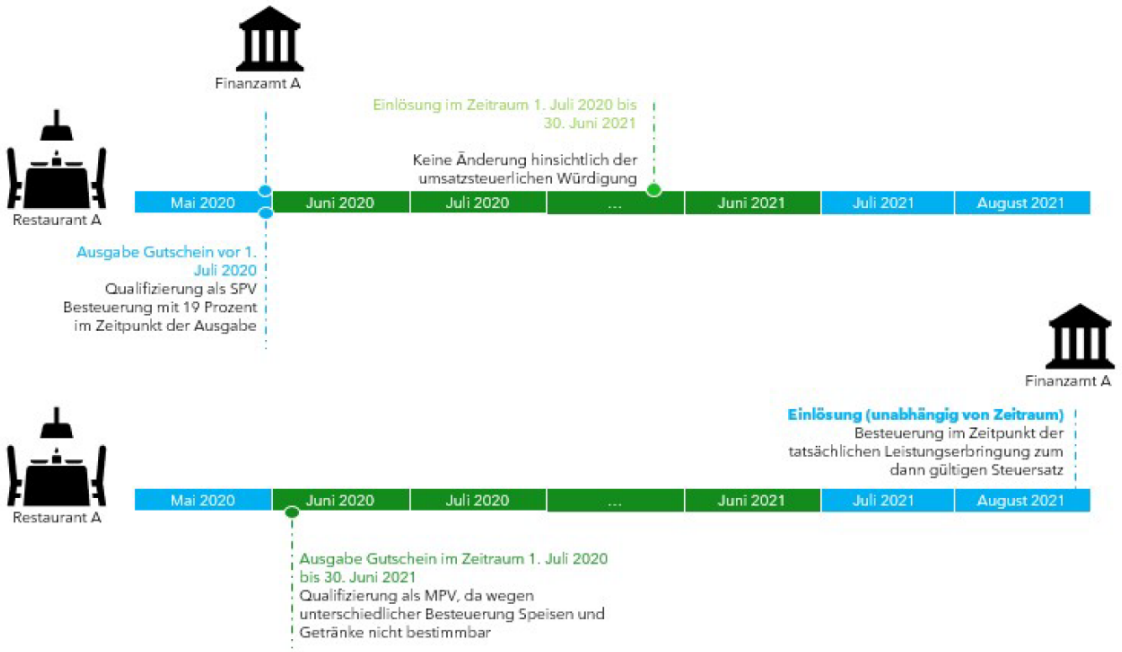

Abbildung 12: Würdigung von Restaurantgutscheinen gemäß BMF-Schreiben vom 4.11.2021.
Eigene Darstellung.

(z.B. 50 EUR) und dieser dann für einen Gegenstand mit weitaus höherem Wert nach dem 31.12.2020 eingesetzt (z.B. 5.000 EUR), so ist die Zuzahlung dem gesenkten Steuersatz ebenfalls zu unterwerfen. Im Hinblick auf die allgemeine Verjährungsvorschrift nach § 195 BGB könnte folglich die temporäre Steuersatzsenkung im Zusammenhang mit einem Einzweck-Gutschein bis 31.12.2023 genutzt werden. Diese Herangehensweise könnte der Finanzverwaltung missfallen, da hier die Zuzahlung nicht mehr im Verhältnis zu der fiktiv ausgeführten Leistung steht. Ob und wie weit diese Gestaltungsmöglichkeit in der Praxis tatsächlich in Anspruch genommen wird ist unklar. Wie eingangs erwähnt ist der Einsatz von Gutscheinen Großteiles im Einzelhandel bzw. Privatkundenbereich anzutreffen. Ein potenzieller Gestaltungsschaden im hohen Maße könnte sich daher m.E. in Grenzen halten. In den wenigsten Fällen dürfte den Konsumenten die Klassifizierung ihres gekauften Gutscheins als auch deren umsatzsteuerliche Würdigung in diesem Ausmaß bekannt sein, so dass dieser daraus eine dementsprechende Steuergestaltung erkennt.

Es ist jedoch nicht von der Hand zu weisen, dass die fiktive Leistung für den Gutschein und deren Gegenleistung (Entgelt) mit einer derartigen Zuzahlung im genannten Beispiel unverhältnismäßig ist. Dahingehend wäre eine Überlegung, ob ab einem bestimmten Zuzahlungsbetrag dieser sich tatsächlich auf einen neuen Umsatz i.S.d. § 3 UStG bezieht und dieser in der Folge dem aktuell geltenden Steuersatz zu unterwerfen ist. Jedoch könnte diese Überlegung zum einen schwer im Einklang mit dem Umsatzsteuergesetz und den dort verankerten Grundsätzen zur Besteuerung von Gutscheinen stehen, und zum andern im Hinblick auf die Umsetzung in der Praxis als nicht praktikabel erweisen. Insgesamt ist festzuhalten, dass die auftretenden Unklarheiten weitreichende Auswirkungen haben, da die Kassensysteme und ERP-Systeme entweder der Verwaltungsauffassung oder der Systematik des Umsatzsteuergesetztes angepasst werden müssen.

4.4. Umsatzsteuerliche Behandlung von COVID-19-Gutscheinen

4.4.1. Ausgangssituation

Das COVID-19-Virus dominiert nach über einem Jahr immer noch die gesamte Weltwirtschaft. Auch wenn in einzelnen Ländern bereits wieder Kulturveranstaltungen stattfinden konnten, so ist dies aktuell in Deutschland noch nicht absehbar. Insbesondere die Reise- und Veranstaltungsbranche wurde dabei schwer getroffen, da eine Leistungserbringung kaum möglich ist und es auch keinen Nachholeffekt geben wird. Um drohenden Insolvenzen dieser Branchen entgegenzuwirken, wurde von der Bundesregierung zwei Gesetz verabschiedet, welche die Ausgabe eines Gutscheins anstelle einer Rückerstattung legitimieren soll.[158] Im Folgenden wird aus Vereinfachungsgründen lediglich auf die Veranstaltungsbranche eingegangen, ein dementsprechendes Pendant findet sich in der Reisebranche.

Voraussetzung zur Anwendung des Gesetzes nach Art. 240 § 5 Abs. 1 EGBGB ist, dass die Eintrittskarte vor dem

[158] Vgl. „Gesetz zur Abmilderung der Folgen der COVID-19-Pandemie im Pauschalreisevertragsrecht und zur Sicherstellung der Funktionsfähigkeit der Kammern im Bereich der Bundesrechtsanwaltsordnung, der Bundesnotarordnung, der Wirtschaftsprüferordnung und des Steuerberatungsgesetzes während der COVID-19-Pandemie"; „Gesetz zur Abmilderung der Folgen der COVID-19-Pandemie im Veranstaltungsrecht und im Recht der Europäischen Gesellschaft (SE) und der Europäischen Genossenschaft (SCE)".

08.03.2020 erworben wurde und die Veranstaltung aufgrund der COVID-19-Pandemie nicht stattfinden kann oder konnte. Dabei hat der Wert des Gutscheins den gesamten Eintrittspreis einschließlich etwaiger Vorverkaufsgebühren zu enthalten. Wird der Gutschein jedoch bis Ende des Jahres 2021 nicht eingelöst, so lebt der Gelderstattungsanspruch nach Art. 240 § 5 Abs. 5 EGBGB wieder auf.

4.4.2. Umsatzsteuerliche Würdigung

Der Veranstalter hat als leistender Unternehmer seinem Kunden mit dem Verkauf einer Eintrittskarte eine bestimmte Leistung zu einem bestimmten Zeitpunkt versprochen. Hierbei handelt es sich um eine Anzahlung nach § 13 Abs. 1 Nr. 1 Buchst. a S. 4 UStG, da das Ticket in der Regel nur eine Art Zahlungsbescheinigung darstellt und die in der Zukunft liegende Veranstaltung die erworbene Leistung ist. Anzahlungsrechnungen unterliegen der Ist-Besteuerung, d.h. die Steuer ist in dem Voranmeldungszeitraum abzuführen, in dem das Entgelt vereinnahmt wurde.

Mit Ausfall der Veranstaltung entfällt die Leistung zur Gegenleistung. Aufgrund eines mangelnde Leistungsaustauschs wird grundsätzlich keine Umsatzsteuer geschuldet, da § 1 Abs. 1 Nr. 1 UStG nicht einschlägig ist. Folglich ist die Bemessungsgrundlage aufgrund von Unmöglichkeit der Leistung nach § 17 Abs. 2 Nr. 2 UStG zu ändern. Grundsätzlich ist diese Vorschrift rein deklaratorisch anzusehen, da aufgrund des fehlenden Umsatzes auch kein Aufwand des Zahlenden zu besteuern ist.[159] Zur Anwendung der Änderung der Bemessungsgrundlage nach § 17 Abs. 2 Nr. 2 UStG wird seitens BFH[160] und Finanzverwaltung[161] weiter die tatsächliche Rückzahlung der geleisteten Anzahlung als Voraussetzung für die Berichtigung des Anzahlungsempfängers gefordert. Fraglich ist, ob durch die Ausgabe des Gutscheins statt der tatsächlichen Rückzahlung diese Voraussetzung als erfüllt gilt. Der Wille des Gesetzgebers ist durch die Gutscheinausgabe diesen als alternatives Mittel der Rückzahlung für den Veranstaltungspreis anzusetzen. Daher könnte dieser Gedanke auch als Voraussetzung der Rückzahlung im Sinne von § 17 UStG übertragen werden.[162] Weiter wird dieser Gedanke dadurch gestützt, dass der Gutschein nach A 3.17 Abs. 1 UStAE als Instrument angesehen wird, der ganz oder teilweise anstelle einer Geldzahlung verwendet werden kann und somit ein äquivalentes Zahlungsmittel darstellt.[163] In der Konsequenz dürften die Voraussetzungen zur Korrektur der Bemessungsgrundlage nach § 17 Abs. 2 Nr. 2 UStG vorliegen und der Veranstalter stünde ein Erstattungsanspruch der rechtsgrundlos abgeführten Umsatzsteuer gegenüber dem Finanzamt zu. Das Pendant bildet ein Rückforderungsanspruch des Finanzamtes gegenüber dem Leistungsempfänger, sofern dieser Unternehmer nach § 2 UStG ist, hinsichtlich der rechtsgrundlos in Abzug gebrachten Vorsteuer.[164]

Im Zeitpunkt der Unmöglichkeit wird der Gutschein als Ersatzinstrument ausgegeben, d.h. im Zeitpunkt der Korrektur entsteht parallel eine mögliche zweite Leistungsbeziehung. Diese Leistungsbeziehung ist abhängig von der Art des ausgegebenen Gutscheins.[165] Gibt der Veranstalter statt des Rückzahlungsbetrags einen Einzweck-Gutschein i.S.v. § 3 Abs. 14 UStG aus, wird ein steuerbarer und steuerpflichtiger Umsatz generiert. D.h. wirtschaftlich betrachtet wird durch die Ausgabe eines Einzweck-Gutscheins keine Liquidität aus der Umsatzsteuererstattung ermöglicht. Der Veranstalter kompensiert die Erstattung der Änderung der Bemessungsgrundlage im gleichen Voranmeldungszeitraum mit der abzuführenden Umsatzsteuer aus der Ausgabe des Einzweck-Gutscheins. Diese Konstellation würde sich in der Folge auch für den Leistungsempfänger liquiditätsneutral gestalten.

Anders ist dies jedoch, wenn der Veranstalter die Voraussetzungen zur Ausgabe eines Mehrzweck-Gutscheins i.S.v. § 3 Abs. 15 UStG erfüllen kann. Die Umsatzsteuer entsteht erst im Zeitpunkt der der tatsächlichen Leistungserbringung, da erst hier der Ort der Leistung und die geschuldete Steuer bekannt sind. Demnach ist im Ergebnis hier eine Möglichkeit für den Veranstalter gegeben einen Liquiditätsvorteil aus dem Erstattungsanspruch bezüglich der bereits abgeführten Umsatzsteuer des ursprünglich geplanten Leistungsaustauschs zu schaffen, ohne durch Ausgabe des Mehrzweck-Gutscheins einen zweiten steuerbaren Umsatz zu begründen. Wird dieser beim Veranstalter gewonnen Liquiditätsvorteil durch das beschriebene Szenario spiegelbildlich beim Leistungsempfänger betrachtet, so erhält dieser weder die vertraglich versprochene Leistung noch sein gezahltes Entgelt zurück. Zudem verliert über die Berichtigung des § 17 Abs. 2 Nr. 2 UStG sein Recht auf Vorsteuerabzug und muss diesen Betrag an das Finanzamt zurückzahlen.[166]

Gemäß Art. 240 § 5 Abs. 5 EGBGB ist dem Leistungsempfänger gestattet sich den Wert des Gutscheins auszahlen zu lassen, sofern dieser nicht bis zum 31.12.2021 eingelöst wurde. Wurde ein Einzweck-Gutschein ausgegeben, so ergeben sich keine umsatzsteuerrechtlichen Konsequenzen gemäß A 3.17 Abs. 7 UStAE, da die Leistung fiktiv als erbracht gilt und diese bereits im Zeitpunkt der Ausgabe des Gutscheins zu versteuern war. Eine Korrekturvorschrift ist hier nicht anwendbar, auch wenn die tatsächliche Leistung nicht erbracht wurde.[167] Weiter ist auch die Nichteinlösung eines Mehrzweck-Gutscheins nach A 3.17 Abs. 13 UStAE umsatzsteuerlich nicht von Bedeutung.

[159] Vgl. Holger Stadie, „§ 17 UStG." In Rau/Dürrwächter UStG (Köln: Dr. Otto Schmidt), Rn. 436.

[160] Vgl. BFH, „Urt. v. 2.9.2010 - V R 34/09." DStR (2010): 2632-2634, Rz. 21; BFH, „Urt. v. 15.9.2011 - V R 36/09." DStR (2011): 2392-2396, Rn. 29.

[161] Vgl. A 17.1 Abs. 7 S. 3 UStAE

[162] Vgl. Rainald Vobbe und Timm Stelzer, „Gutscheine als Liquiditätsquelle: Umsatzsteuerrechtliche Auswirkungen von COVID-19 auf die Reiseverkehrs- und Veranstaltungsbranche." DB, Nr. 21 (2020): 1086– 1092, S. 1088.

[163] Vgl. Huschens, „Einzweck- und Mehrzweck-Gutscheine in der USt - Praktische Auswirkungen der ab 01.01.2019 geltenden Neuregelungen", S. 450.

[164] Vgl. Holger Stadie, „§ 17 UStG." In Rau/Dürrwächter UStG, Rn. 434.

[165] Vgl. Stefanie Becker, „Ausgabe von "Corona-Gutscheinen": Umsatzsteuerliche Folgen." BC (2020): 202- 203, S. 2020.

[166] Vgl. Vobbe und Stelzer, „Gutscheine als Liquiditätsquelle", S. 1089.

[167] Vgl. Axel Leonard und Martin Robisch, „§ 3 UStG." In Bunjes UStG, Rn. 330. Rudolf Hahn, „§ 3 UStG." In BeckOK UStG, Rn. 610.

Jedoch stellt sich berechtigterweise die Frage, inwieweit sich dies ändern könnte, da nach dem 31.12.2021 der Erstattungsanspruch in Geld des Leistungsempfängers gegenüber dem leistenden Unternehmer auflebt, d.h. der generierte Umsatz rückabgewickelt wird. Da lediglich bei der Ausgabe eines Einzweck-Gutscheins ein steuerbarer und steuerpflichtiger Umsatz generiert wurde, ist auch nur in diesem Fall beim leistenden Unternehmer die Bemessungsgrundlage nach § 17 Abs. 2 Nr. 3 UStG zu ändern.[168]

4.5. Umsatzsteuerliche Behandlung von Städte-Cards unter Berücksichtigung der Gutschein-Richtlinie

Das BMF hat in A 3.17 Abs. 1 UStAE hat eine Vielzahl von Sachverhalten aufgeführt, welche von der Gutscheindefinition als negativ abzugrenzen sind. Unklar ist die Frage, ob bei es sich bei sog. Städte-Cards um einen Gutschein i.S.d. Richtlinie handeln könnte.[169]

Gästekarten sind ein beliebtes Instrument zur Tourismusförderung, welche in der Regel ihren Inhabern das Recht verleihen eine Vielzahl touristischer Attraktionen in einer Stadt (z.B. Eintritt in Museen, Zoos, etc.) in Anspruch zu nehmen. Ergänzt werden diese oftmals durch Rabattcoupons für Restaurants oder den Einzelhandel und um die Nutzung des öffentlichen Nahverkehrs.[170] Die am Städtekartenmodell teilnehmenden Unternehmen profitieren dabei von einer positiven Auswirkung auf ihren Bekanntheitsgrad, während der einzelne Tourist finanzielle Vorteile bezieht. Zusammengefasst bündeln Städtekarten eine mehrere Funktionen: Marketinginstrument, Zahlkarte, Rabattinstrument und ggf. auch Gutschein, woraus jeweils unterschiedliche Rechtsfolgen resultieren könnten.[171] Da die Ausgestaltung bzw. das Geschäftsmodell zur Abgabe einer Städtekarte von Stadt zu Stadt unterschiedlich ist, kann eine grundsätzliche Aussage zur umsatzsteuerlichen Würdigung nicht getroffen werden.

Jedoch wird in der Praxis regelmäßig das sog. Bezahlmodell angewendet. Dieses beinhaltet den Verkauf der Gästekarte von einem Systembetreiber direkt an den Touristen.[172] Dagegen wird beim Gastgebermodell vom Touristen für die Gästekarte kein Entgelt geleistet, sondern wird diese im Rahmen des Beherbergungsvertrags ausgehändigt. Der Systembetreiber erhält hierfür eine Umlage vom Gastgeber.[173] Unter Beachtung verschiedener Faktoren errechnet sich die jeweilige Erlösbeteiligung der Leistungsträger.[174]

Da regelmäßig das Gastgebermodell lediglich zu Preisnachlässen berechtigt, soll im Rahmen dieser Arbeit der Fokus auf der umsatzsteuerlichen Würdigung im Zusammenhang mit der Gutscheinrichtlinie auf dem Bezahlmodell liegen.

Bei Ausgabe der Städtekarte sind zunächst die Leistungsbeziehungen zu identifizieren, da der Systembetreiber in der Regel nicht der Leistungserbringer ist. Der Großteil der Einnahmen werden aufgrund der Zahlungen an die Leistungserbringer nicht beim Systembetreiber verbleiben. Daher ist die Überlegung, ob der Systembetreiber lediglich die Funktion einer Mittelsperson gemäß A 10.4 Abs. 1 S.2 UStAE zuteil wird und die Einnahmen lediglich als durchlaufender Posten gemäß § 10 S. 5 UStG zu klassifizieren sind. Durchlaufende Posten sind Beträge, die der Unternehmer im Namen und für Rechnung eines anderen vereinnahmt und verausgabt. Laufende Rechtsprechung setzt zur Qualifizierung als durchlaufender Posten voraus, „dass unmittelbare Rechtsbeziehungen zwischen zwei Beteiligten bestehen, in die der Unternehmer nur als vermittelnde Person, wie eine Zahlstelle, zwischengeschaltet ist.“[176] Diese Systematik wird in der Regel bei Städtekarten nicht der Fall sein, da die Touristen sich unmittelbar dem Systembetreiber zu einer Zahlung verpflichten. Grundsätzlich ist davon auszugehen, dass ein Handeln im eigenen Laden auch auf eigenem Namen geschieht. Von einer Vermittlereigenschaft ist nur dann anzunehmen, wenn dem Kunden von vornherein erkennbar ist, dass der Ladeninhaber nicht als Eigenhändler auftritt.[177] Ist demzufolge eine positive Rechtsbeziehung zwischen Systembetreiber und dem Touristen zu bestätigen, ist die Frage zu klären, ob es sich bei der Ausgabe der Städtekarte um einen Mehrzweck-Gutschein handeln könnte.[178]

Interessant ist, dass bereits 2019 BusinessEurope die Frage hinsichtlich der umsatzsteuerlichen Einordnung von Gästekarten im Bezahlmodell an den Mehrwertsteuerausschuss adressierte.[179] Die Autoren der Anfragen führten mehrere Alternativen zur umsatzsteuerlichen Einordnung und Besteuerung auf:

(1) Klassifizierung als Mehrzweck-Gutschein
(2) Klassifizierung als Mehrzweck-Gutschein mit zusätzlicher Margensteuer
(3) Klassifizierung als Leistungskette

Die Tabelle 2 zeigt, dass bei der Klassifizierung der Städtekarte als Mehrzweck-Gutschein die geringste Umsatzsteuerschuld entstehen würde. Die Ausgabe des Mehrzweck- Gut-

[168] Vgl. Suse Martin, „§ 3 UStG.” In Sölch/Ringleb Umsatzsteuergesetz, 89. EL Juni 2020:Rn. 811.
[169] Alena Kirchinger und Maximilian Dachauer, „Städte-Cards als Gutscheine i.S.d. Gutschein-Richtlinie.“ ustb (2021): 22–26, S. 23.
[170] Vgl. Christian Sterzinger, „Umsatzsteuerliche Behandlung von touristischen Gästekarten-Modellen.“ ustb (2020): 94–100, S. 23.
[171] Vgl. Hans-Martin Grambeck, „Gäste-Karten - Versuch einer umsatzsteuerlichen Einordnung.“ MwStR (2019): 940-942, S. 940.
[172] Vgl. ebd.
[173] Vgl. Grambeck, „Gäste-Karten - Versuch einer umsatzsteuerlichen Einordnung“, S. 942.
[174] Vgl. ebd., S. 940.
[175] In Anlehnung an: ebd.
[176] BFH, „Urt. v. 04.05.2011 - XI R 4/09.“ BeckRS (2011): 96066, Rz. 17.
[177] Vgl. Martin Robisch, „§ 1 UStG.“ In Bunjes UStG (München: C.H. Beck oHG), Rz. 98.
[178] Vgl. Kirchinger und Dachauer, „Städte-Cards als Gutscheine i.S.d. Gutschein-Richtlinie“, S. 24.
[179] Vgl. VAT Committee, „Information Paper v. 3.4.2019: VAT treatment of ”city cards”.“. https://circabc.europa.eu/ui/group/cb1eaff7-eedd-413d-ab88-94f761f9773b/library/20546cf6-42a6-4196-b14f-d875b3826c06/details (letzter Zugriff: 28. April 2021), S. 2.

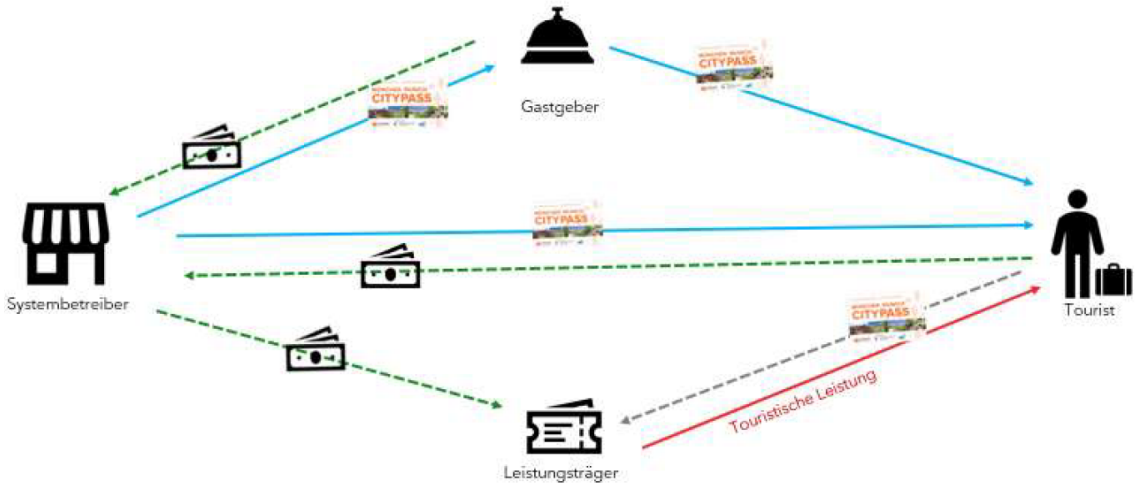

Abbildung 13: Geschäftsmodelle von Gästekarten.[175]

Tabelle 2: Mögliche umsatzsteuerliche Einordnung von Städtekarten.
Vgl. VAT Committee, „Information Paper v. 3.4.2019: VAT treatment of "city cards".". https://circabc.europa.eu/ui/group/cb1eaff7-eedd-413d-ab88-94f761f9773b/library/20546cf6-42a6-4196-b14f-d875b3826c06/details (letzter Zugriff: 28. April 2021), S.5-7

		Mehrzweck- Gutschein	Mehrzweck- Gutschein mit zusätzlicher Margensteuer	Leistungskette
Ausgabe der Städtekarte	BMG (in EUR)	500	150	420,17
	Steuersatz	0%	19%	19%
	USt (in EUR)	0	28,5	79,83
	(Brutto)Preis (in EUR)	500	500	500
Einsatz der Städtekarte		Öffentlicher Nahverkehr §12 Abs. 2 Nr.10 UStG	Eintritt Freizeitpark § 12 Abs. 1 UStG	Eintritt Museum § 4 Nr. 20 Buchst. a UStG
	BMG (in EUR)	93,46	84,03	100
	Steuersatz	7%	19%	0%
	USt (in EUR)	6,54	15,97	0
	(Brutto)Preis (in EUR)	100	100	100
Summe USt		22,51	51,01	102,34

scheins unterliegt gemäß § 3 Abs. 15 UStG aufgrund der nicht feststehenden Steuerschuld erst bei Einlösung der Umsatzsteuer. Wird die Städtekarte für den öffentlichen Nahverkehr und den Eintritt in den Freizeitpark und in das Museum genutzt, kann die Umsatzsteuer in Bezug auf die tatsächlich erbrachte Steuer berechnet werden. Bei der Klassifizierung als Mehrzweck-Gutschein würde sich eine gesamte Steuerbelastung in Höhe von (6,54 + 15,97=) EUR 22,51 ergeben. Bei der Annahme, dass ein Mehrzweck- Gutschein mit einer zusätzlichen Margenbesteuerung vorliegt, ergibt sich im Beispiel ein Umsatzsteuerbetrag in Höhe von EUR 51,01. Dabei wird in diesem Beispiel davon ausgegangen, dass die Marge EUR 150,00 beträgt. Diese berechnet sich aus der Differenz des Ausgabepreises von EUR 500,00 abzüglich der Leistungen (EUR 300,00) und einem Verwaltungsaufwand von EUR 50,00. Betrachtet man die letzte Alternative, sodass jeder Vorgang eine Leistungserbringung darstellt, so würde eine Umsatzsteuerbelastung in Höhe von EUR 102,34 anfallen. Hier wird bereits davon ausgegangen, dass die Ausgabe der Städtekarte eine Leistungsbeziehung darstellt und diese der Umsatzsteuer unterliegt (USt: EUR 79,83). Hinzu kommt die jeweilige Leistungserbringung aus der Nutzung des öffentlichen Nahverkehrs und den Eintritten (USt: EUR 15,97).

Nach Auffassung des Mehrwertsteuerausschusses ist entscheidend, ob die Städtekarte Merkmale aufweist, um der Definition eines Gutscheins gerecht zu werden. Städtekarte, welche lediglich und ausschließlich zu Preisnachlässen berechtigen, ohne den Zugang zu der Attraktion garantieren, sind keine Gutscheine i.S.d. Gutschein-Richtlinie. Im Übrigen könnten Städtekarten jedoch unter die Gutschein-Richtlinie fallen. Eine Festlegung hinsichtlich eines Einzweck- oder Mehrzweck-Gutscheins ist abhängig von der im Einzelfall abgedeckten Leistungen. Regelmäßig werden neben umsatzsteuerpflichtigen Leistungen auch umsatzsteuerfreie Leistungen durch die Städtekarte ermöglicht. Somit ist im Zeitpunkt der Ausgabe die Bestimmung des Steuersatzes nicht mög-

lich.[180] Gerade im touristischen Bereich ist die Einschlägigkeit der Steuerbefreiung nach § 4 Nr. 20 UStG keine Seltenheit. Die Berechtigung zum Eintritt von Museen, zoologischen Gärten und Theatern ist regemäßig Teil der jeweiligen Städtekarten.[181] Eine zentraler Hinweis des Mehrwertsteuerausschuss ist, dass bei einer Bündelung von Funktionen, wie beispielsweise Eintrittsberechtigung als auch Beförderungsberechtigung, diese nicht zu einer Differenzierung der Leistungselemente führen solle.[182] Insgesamt erscheint die Beurteilung des Mehrwertsteuerausschusses als nachvollziehbar und stimmig. Unabdingbar ist jedoch eine genaue Analyse des jeweiligen Modells bei der Ausgabe als auch der Funktionen der Städtekarte. Ist das Ergebnis der Analyse die Bestätigung eines Mehrzweck-Gutscheins nach § 3 Abs. 15 UStG, so erfüllt die Ausgabe der Städtekarte durch den Systembetreiber lediglich eine Bargeldersatzfunktion und es liegt ein umsatzsteuerlich irrelevanter Tausch Geld gegen elektronisches Geld vor.[183]

Wird die Städtekarte vom Touristen nicht eingelöst, so könnte die Vorgabe des A 3.17 Abs. 13 S. 2 UStAE zu berücksichtigen sein. Das BMF geht nach A 3.17 Abs. 10 UStAE grundsätzlich bei Mehrzweck-Gutscheinen in Vertriebsketten von einer steuerbaren Leistung des ausstellenden Unternehmers an den leistenden Unternehmer aus. Daher ist bei einem Nichteinsatz der Städtekarte davon auszugehen, dass der Systembetreiber aus dem vollen Kaufpreis die Umsatzsteuer anzumelden und abzuführen hat.

Da Städtekarten eine hohe Popularitätsrate genießen, ist es umso erstaunlicher, dass deren umsatzsteuerliche Behandlung in der Fachliteratur und Rechtsprechung kaum zur Geltung kommen und nicht abschließend geklärt sind. Eine Klassifizierung der Städtekarten als Mehrzweck-Gutscheine ist m.E. plausibel, jedoch ist das einzelne Modell zu analysieren. Sofern eine Eingruppierung als Gutschein zweifelhaft erscheint, ist eine Abstimmung mit dem jeweiligen Finanzamt durch eine verbindliche Auskunft i.S.d. § 89 AO empfehlenswert.

5. Thesenförmige Zusammenfassung

Im Folgenden werden die Ergebnisse dieser Arbeit thesenförmig zusammengefasst:

- Gutscheine jeglicher Art werden in mannigfaltigen Branchen verwendet. Insbesondere im Einzelhandel ist der Gutschein ein etabliertes Instrument zur Neukundengewinnung bzw. Stammkundenpflege sowie Absatzförderung. Die Vielzahl der Gutscheine differenziert sich aus betriebswirtschaftlicher und rechtlicher Sicht über den Verpflichtungsgrad des ausgebenden Unternehmers gegenüber des Gutscheininhabers. Insbesondere die Möglichkeit mit der Ausgabe von Gutscheinen liquide Mittel, ohne sofortige Gegenleistung, zu generieren wurde insbesondere im Zusammenhang mit der Corona-Krise für Unternehmen, die Öffentlichkeit und den Gesetzgeber deutlich.

- Die umsatzsteuerliche Würdigung von Gutscheinen war bis zur Umsetzung der Gutschein-Richtlinie zum 1.1.2019 nicht im nationalen UStG verankert und war in der Vergangenheit geprägt von Rechtsunsicherheiten, vornehmlich in komplexeren Vertriebsketten und/oder grenzüberschreitenden Transaktionen. Eine fehlende Harmonisierung innerhalb der EU-Mitgliedstaaten führten dabei zu Missverhältnissen der Besteuerung. Diese sollten mit der sog. Gutschein-Richtlinie innerhalb der EU eliminiert werden.

- Mit BMF-Schreiben vom 2.11.2020 erfolgte eine Präzisierung hinsichtlich zahlreicher offener Fragestellungen zur Klassifizierung des jeweiligen Gutscheins als auch zur umsatzsteuerlichen Würdigung von Gutscheinen. Die Analyse des BMF-Schreibens zeigt jedoch, dass die zusätzlich von der Finanzverwaltung geforderten Abgrenzungskriterien als Einzweck- bzw. Mehrzweck-Gutscheine durchaus weitere Rechtsunsicherheit mit sich bringen, statt diese zu beheben. Zusammenfassend werden hierbei teilweise vom Gesetz abweichende Anforderungen an die Qualifizierung als Einzweck- bzw. Mehrzweck-Gutscheins gestellt. Diese nationalen Vorgaben scheinen dem gesetzten Ziel einer einheitlichen unionsweiten Lösung nicht gerecht zu werden.

- Die umsatzsteuerliche Würdigung von Einzweck- und Mehrzweck-Gutscheine außerhalb von Vertriebsketten ist dem Grunde nach der bisherigen Würdigung von sog. (Nenn-)Wertgutscheinen und Waren- oder Sachgutscheinen identisch. Das Vorliegen eines Einzweck-Gutscheins fingiert eine steuerbare und steuerpflichtige Lieferung im Zeitpunkt der Ausgabe des Gutscheins. Dagegen wird der Mehrzweck-Gutschein vielmehr als Zahlungsmittel angesehen, der erst im Zeitpunkt der Einlösung der Besteuerung unterliegt. Somit darf es kaum verwundern, dass durch die Analyse insbesondere offene Fragestellungen im Zusammenhang mit Einzweck-Gutscheinen identifiziert wurden. Hervorzuheben sei hier die Ortsbestimmung bei Lieferungen in Zusammenhang mit Einzweck-Gutscheinen. Die Finanzverwaltung nimmt für die fiktive Lieferung eine ruhende Lieferung nach § 3 Abs. 7 UStG an. Diese Ansicht ist jedoch nicht im Gesetz verankert und könnte sich von möglicherweise von anderen Ländern abweichen. Sollte mit Einlösung des Gutscheins grenzüberschreitende Transaktionen vorliegen, könnte es folglich

[180] Vgl. VAT Committee, „Working paper No 983 v. 13.11.2019: Questions raised following implementation of the Voucher Directive.“. https://circabc.europa.eu/ui/group/cb1eaff7-eedd-413d-ab88-94f761f9773b/library/8ded663d-0886-4d3a-b036-2d8c77e5d66a/details (letzter Zugriff: 28. April 2021), S. 24.

[181] Vgl. Kirchinger und Dachauer, „Städte-Cards als Gutscheine i.S.d. Gutschein-Richtlinie“, S. 24.

[182] Vgl. VAT Committee, „Working paper No 983 v. 13.11.2019“, S. 26.

[183] Vgl. Sterzinger, „Umsatzsteuerliche Behandlung von touristischen Gästekarten-Modellen“, S. 99.

wieder zu Diskrepanzen zwischen den einzelnen EU-Mitgliedstaaten kommen.

- Das BMF-Schreiben hatte sich weiter zur umsatzsteuerlichen Würdigung von Einzweck- und Mehrzweck-Gutscheinen innerhalb von Vertriebsketten positioniert. Schwächen zeigen sich hier vor allem durch die geforderte Meldepflicht in Zusammenhang mit Einzweck-Gutscheinen in Vertriebsketten (Handeln im eigenen Namen). Es entsteht eine Abhängigkeit des leistenden Unternehmers zum ausstellenden Unternehmer hinsichtlich der eigenen Steuerschuld.

- Die temporäre Steuersatzsenkung im 2. Halbjahr hatte Einfluss auf die Würdigung der Gutscheine. Hier konnten sich über den Kauf von Gutscheinen ein legales Gestaltungspotential entwickeln, um sich den geminderten Steuersatz zu sichern. Diese Möglichkeit wurde jedoch seitens der Finanzverwaltung deutlich, jedoch ohne weitere Begründung, abgewehrt.

- Es scheint dem BMF insbesondere ein Anliegen gewesen zu sein, die umsatzsteuerliche Würdigung von Gutscheinen im Zusammenhang mit Restaurantdienstleistungen und dem weiterhin abgesenkten Steuersatz für die Abgabe von Speisen darzulegen. Jedoch dürften diese Ausführungen für die meisten Gastronomen dahingehend obsolet sein, da in Kombination mit dem seit der Corona- Krise etabliertem „To-Go-Geschäft“ dem Grunde nach regelmäßig ein Mehrzweck- Gutschein vorliegen wird.

- Weiter wurde im Zusammenhang mit der Steuersatzsenkung eine vom Gesetz abweichende Beurteilung in Bezug auf Zuzahlungen bei Einzweck-Gutscheinen seitens der Finanzverwaltung dargelegt. Würde man der Auffassung des BMF folgen, so wäre davon auszugehen, dass eine Zuzahlung als eigenständige Leistung zu beurteilen ist. Dieser Ansatz widerspricht jedoch § 17 UStG. Jedoch ist es nicht abzustreiten, dass bei Anwendung des § 17 UStG eine Gestaltungsmöglichkeit gegeben ist, die im gewissen Rahmen in dieser Art und Weise nicht vom Gesetzgeber vorgesehen wurde. Eine dahingehende Rechtssicherheit ist für die Praxis eminent.

- Im Rahmen der Corona-Pandemie wurde seitens der Bundesregierung der Vorteil der Gutscheine erkannt und ein Gesetz für die Veranstaltungs- und Reisebranche verabschiedet, welche die Ausgabe eines Gutscheins statt der Rückerstattung in Geld legitimieren soll. Die umsatzsteuerliche Würdigung der gesamten Leistungskette hat gezeigt, dass sich der leistende Unternehmer weitere liquide Vorteile sichern kann, sofern der ausgegebene Gutschein als Mehrzweck-Gutschein klassifiziert werden kann.

- Sogenannte Städtekarten wurde im Rahmen der Definition des BMFs nicht berücksichtigt. Dabei wurde bereits 2019 die Frage nach der umsatzsteuerlichen Einordnung der Städtekarten an den Mehrwertsteuerausschuss adressiert. Im Ergebnis kann eine Städtekarte als Mehrzweck-Gutschein klassifiziert werden, wenn diese die geforderten Merkmale aufweist. Unabdingbar ist jedoch eine genaue Analyse des dahinterstehenden Geschäftsmodells. Eine verbindliche Auskunft i.S.d. § 89 AO beim zuständigen Finanzamt dürfte im Regelfall weiterhin empfehlenswert sein.

- Positiv zu werten ist, dass eine unionsübergreifende Regelung zur umsatzsteuerlichen Würdigung von Gutscheinen etabliert werden konnte. Die Rechtsunsicherheiten in Bezug auf die Klassifizierung als auch zur Würdigung bleiben dennoch bestehen. Im Ergebnis werden sich einige der genannten Fragen erst in den kommenden Jahren vor Gericht lösen können. Unternehmen, die Gutscheine ausgeben haben diese erneut zu würdigen und unter der seit 1.1.2019 geltenden Gesetzeslage als auch dem veröffentlichten BMF-Schreiben vom 2.11.2020 umsatzsteuerlich zu hinterfragen.

Literatur

Ahrens, M. (1996). Gutscheine. *Betriebs-Berater*, 2477–2481.

Bader, J., Meissner, G., Bobikov, M., Wahl, A., Ravensberg, S. & Schempp, S. (2020). *Umsatzsteuerrecht visualisiert: Das deutsche Umsatzsteuerrecht in Übersichten* (3. Aufl.). Stuttgart, Schäffer-Poeschel Verlag.

Becker, S. (2020). Ausgabe von "Corona-Gutscheinen": Umsatzsteuerliche Folgen. *Zeitschrift für Bilanzierung, Rechnungswesen & Controlling*, 202–203.

Becker, S. (2021). Behandlung von Gutscheinen im Umsatzsteuerrecht: Offene Fragen nach dem BMF-Schreiben v. 2.11.2020 sowie Auswirkungen für die Praxis. *Mehrwertsteuerrecht*, 223–229.

Boost Internet GmbH. (2021). *Erlebnisgeschenke-Report: Eine Studie zum Markt für Erlebnisgeschenke in Deutschland.* http://www.erlebnisgeschenke.de/download/marktstudie.pdf. (Zuletzt abgerufen am: 03.03.2021)

Bundessteuerberaterkammer. (2021a). *BMF-Schreiben zur Anwendung der Neuregelungen zu den Gutscheinen.* https://www.bstbk.de/downloads/bstbk/presse-und-kommunikation/stellungnahmen/BStBK_2019-006_Stellungn_USt_Gutscheine.pdf. (Zuletzt abgerufen am: 25.01.2021)

Bundessteuerberaterkammer. (2021b). *Diskussionsentwurf für eine gesetzliche Umsetzung der Gutschein-Richtlinie (EU) 2016/1065.* https://www.bstbk.de/downloads/bstbk/presse-und-kommunikation/stellungnahmen/BStBK_2017-019_Diskussionsentwurf-gesetzliche-Umsetzung-Gutschein_1.pdf. (Zuletzt abgerufen am: 01.02.2021)

Bunjes, J. & Geist, R. (2020). *Bunjes UStG* (19. Aufl.). München, C.H. Beck.

Burmann, C., Gunnar, C., Esch, F.-R., Henning, A., Kenning, P., Kirchgeorg, M. & Markgraf, D. e. a. (2013). *Kompakt-Lexikon Marketingpraxis: 2.200 Begriffe nachschlagen, verstehen, anwenden*. Wiesbaden, Springer Gabler.

Deloitte GmbH Wirtschaftsprüfungsgesellschaft. (2017). *Christmas Survey 2017: Zentrale Ergebnisse für Deutschland.* https://www2.deloitte.com/de/de/pages/consumer-business/articles/christmas-survey.html. (Zuletzt abgerufen am: 24.02.2021)

Deloitte GmbH Wirtschaftsprüfungsgesellschaft. (2018). *Christmas Survey 2018: Zentrale Ergebnisse für Deutschland.* https://www2.deloitte.com/de/de/pages/consumer-business/articles/christmas-survey.html. (Zuletzt abgerufen am: 24.02.2021)

Deloitte GmbH Wirtschaftsprüfungsgesellschaft. (2019). *Christmas Survey 2019: Zentrale Ergebnisse für Deutschland.* https://www2.deloitte.com/de/de/pages/consumer-business/articles/christmas-survey.html. (Zuletzt abgerufen am: 24.02.2021)

Deloitte GmbH Wirtschaftsprüfungsgesellschaft. (2020). *Christmas Survey 2020: COVID-19 halbiert das deutsche Weihnachtsbudget.* https://www2.deloitte.com/de/de/pages/consumer-business/articles/christmas-survey.html. (Zuletzt abgerufen am: 24.02.2021)

Denker, D. & Schürkötter, N. (2020). Umsatzbesteuerung von Gutscheinen in der Praxis: Wieviel Klarheit bringt das BMF-Schreiben vom 2.11.2020. *USt direkt digital* (23), 14–19.

Deutscher Industrie- und Handelskammertag e.V., Zentralverband des deutschen Handwerks e.V., Bundesverband deutscher Banken e.V., Handelsverband Deutschland (HDE), Bundesverband der deutschen Industrie e.V., Bundesvereinigung der deutschen Arbeitgeberverbände e.V., ... Bundesverband Großhandel, Außenhandel, Dienstleistungen e.V. (2020). *Stellungnahme zum Umsetzungsvorschlag der Gutschein- Richtline (EU) 2016/1065 durch die Finanzverwaltung: Gz: III C 2 - S7100/07/10045:003; DOK 2017/0597914.* https://www.gdv.de/resource/blob/5720/c147f626743d08d287241a9a30a4fc55/zum-umsetzungsvorschlag-der-gutschein-richtlinie-1309327852-data.pdf. (Zuletzt aufgerufen am: 28.12.2020)

Deutscher Industrie- und Handelskammertag e.V., Zentralverband des deutschen Handwerks e.V., Bundesverband deutscher Banken e.V., Handelsverband Deutschland (HDE), Bundesverband der deutschen Industrie e.V., Bundesvereinigung der deutschen Arbeitgeberverbände e.V., ... Bundesverband Großhandel, Außenhandel, Dienstleistungen e.V. (2021). *Stellungnahme zum Entwurf eines BMF- Schreibens zur umsatzsteuerlichen Behandlung von Einzweck- und Mehrzweck- Gutscheinen: GZ: III C 2 - S7100/19/10001:002; DOK 2019/1090742.* https://www.gdv.de/resource/blob/56412/d42d59ae582f10685246106db38d40d0/stellungnahme-gutscheine---download-data.pdf. (Zuletzt abgerufen am: 05.03.2021)

DEUTSCHER STEUERBERATERVERBAND e.V. (2021). *Referentenentwurf eines Jahressteuergesetzes 2018.* https://www.dstv.de/interessenvertretung/steuern/stellungnahmen-steuern/2018-s07-refentwurf-jstg2018. (Zuletzt abgerufen am: 19.05.2021)

Drüen, K.-D. (2021). §42 Abgabenordnung. In *Tipke/Kruse Abgabenordnung - Finanzgerichtsordnung: AO, FGO.* Köln, Verlag Dr. Otto Schmidt.

Duden. (2019). *Duden - Deutsches Universalwörterbuch* (9. Aufl.). Berlin, Dudenverlag.

Ebenroth, C. T., Boujong, K., Joost, D. & Strohn, L. (Hrsg.). (2020). *Handelsgestzbuch Band 1* (4. Aufl.). C.H. Beck/Verlag Vahlen.

Eckert, K.-H. (2020). Die Gutscheinregelung im Zusammenhang mit der Absenkung der Umsatzsteuer. *USt direkt digital* (19), 15–16.

Eilnachricht. (2018). Liquidität: Mit Gutscheinen Zahlungsfähigkeit verbessern. *nwb BB* (9), 260.

Englisch, J. (2017). Das neue MwSt-Sonderregime für Gutscheine. *ifst-Schrift 515*.

Erdbrügger, A. (2020). Die temporäre Senkung der Umsatzsteuersätze - Fluch oder Segen? *Umsatzsteuer Rundschau*, 485–491.

Esprit Global Image GmbH. (2021). *Esprit Fanchise Store Mindelheim.* https://www.esprit.com/storefinder?storeid=13470872&location=48.0459,10.4872. (Zuletzt abgerufen am: 18.03.2021)

Europäische Kommission. (2021a). *COM (2012) 206 final: Vorschlag für eine Richtlinie des Rates zur Änderung der Richtlinie 2006/112/EG über das gemeinsame Mehrwertsteuersystem hinsichtlich der Behandlung von Gutscheinen.* https://ec.europa.eu/transparency/regdoc/rep/1/2012/DE/1-2012-206-DE-F1-1.Pdf. (Zuletzt abgerufen am: 04.02.2021)

Europäische Kommission. (2021b). *Kommission schlägt neue MwSt-Vorschriften für Gutscheine vor.* https://ec.europa.eu/commission/presscorner/detail/de/IP_12_464. (Zuletzt abgerufen am: 21.02.2021)

Feil, K., Kupke, N. & Greisl, E. (2012). Aktuelle Entwicklungen bei der umsatzsteuerlichen Behandlung von Gutscheinen. *Betriebs-Berater*, 3113–3123.

Feil, K. & Polok, V. (2017). Hält die EU-Gutschein-Richtlinie was sie verspricht? *Mehrwertsteuerrecht*, 65–68.

Fink, K.-J. (2007). *888 Weisheiten und Zitate für Finanzprofis*.

Fritsch, F. (2019a). §3 UStG. In *Reiß/Kraeusel/Langer UStG* (Bd. 164. Lieferung). Bonn, Stollfuß.

Fritsch, F. (2019b). §3 UStG. In *Reiß/Kraeusel/Langer UStG* (Bd. 165. Lieferung). Bonn, Stollfuß.

Gerhards, D. (2021). Gutscheine im Umsatzsteuerrecht - BMF-Schreiben vom 2.11.2020. *Deutsche Steuer-Zeitung* (8), 307–318.

GfK. (2021). *Was lag unter dem Weihnachtsbaum? Ein Rückblick auf das Weihnachtsgeschäft 2017.* https://www.gfk.com/de/insights/was-lag-unter-dem-weihnachtsbaum. (Zuletzt abgerufen am: 04.02.2021)

Grambeck, H.-M. (2019). Gäste-Karten - Versuch einer umsatzsteuerlichen Einordnung. *Mehrwertsteuerrecht*, 940–942.

Grube, U. (2019). Der Ton wird rauer. *DATEV magazin*, *26* (12), 9–11.

Grünwald, U. & Pogodd, A. (2001). Ausgabe von Wertgutscheinen durch Händler: Hinweise zur Vermeidung umsatzsteuerlicher Risiken. *Der Umsatz-Steuer-Berater*, 212–215.

Hahn, R. (2019). §3 UStG. In R. Weymüller (Hrsg.), *Beck'scher Online-Kommentar UStG (BeckOK UStG)*. München, C.H. Beck.

Hammerl, R. & Fietz, A. (2020). Umsatzsteuerliche Behandlung von Einzweck- und Mehrzweck-Gutscheinen: BMF klärt mit Schreiben v. 2.11.2020 Anwendungsfragen. *NWB – Steuer- und Wirtschaftsrecht* (48), 3538–3549.

Haric, P. (2021). *Unternehmer: Definition: Was ist "Unternehmer"?* https://wirtschaftslexikon.gabler.de/definition/unternehmer-49340. (Zuletzt abgerufen am: 24.03.2021)

Heermann, P. W. (2020). §783 BGB. In M. e. a. Junker (Hrsg.), *juris Praxiskommentar BGB: Band 2 - Schuldrecht*. Saarbrücken, Juris.

Heinrichshofen, S. & Kupke, N. (2017). Gutscheine nach der Änderung der Mehrwertsteuer-Systemrichtlinie (MwStSystRL) - Anmerkungen zu den Auswirkungen auf die deutsche Besteuerungspraxis: (Teil 1). *Umsatzsteuer- und Verkehrsteuer-Recht*, 142–148.

Henssler, M. (Hrsg.). (2020). *beck-online.GROSSKOMMENTAR*. München, C.H. Beck.

Herrler, H. & Herrler, C. (2019). Mehr Netto vom Brutto. *NWB – Steuer- und Wirtschaftsrecht* (6), 350–364.

Huschens, F. (2019). Einzweck- und Mehrzweck-Gutscheine in der USt - Praktische Auswirkungen der ab 01.01.2019 geltenden Neuregelungen. *DER BETRIEB* (9), 446–456.

Huschens, F. (2020). Einzweck- und Mehrzweck-Gutscheine - Die zu erwartende Rechtsauffassung der Verwaltung. *Umsatzsteuer- und Verkehrsteuer-Recht*, 48–56.

Institut für Wissen in der Wirtschaft GmbH. (2020). Umsatzsteuer: Corona-Erleichterungen für Gastronomie und die Auswirkung auf Gutscheinverkäufe. *Aktuelles aus dem Steuer- und Wirtschaftsrecht*, 473–474.

Junker, M. et al. (Hrsg.). (2020). *juris Praxiskommentar BGB: Band 2 - Schuldrecht* (9. Aufl.). Saarbrücken, Juris.

Kirchinger, A. & Dachauer, M. (2021). Städte-Cards als Gutscheine i.S.d. Gutschein-Richtlinie. *Der Umsatz-Steuer-Berater*, 22–26.

Klein, F. (2020). *Klein AO* (15. Aufl.). München: C.H. Beck.

Knöfel, O. L. (2017). Gift Cards (Geschenkkarten) im Einzelhandel: Rechtsnatur, Funktionsweise, Probleme. *Zeitschrift für Wirtschafts- und Bankrecht* (18), 833–840.

Korn, C. (2020). §17 UStG. In *Bunjes UStG.* München, C.H. Beck.

Koss, C. (2019). Gutscheine und ihre rechtliche Beurteilung. *DER BETRIEB* (47), 2593–2599.

Kraeusel, J. (2019). §14 UStG. In *Reiß/Kraeusel/Langer UStG* (Bd. 165. Lieferung). Bonn, Stollfuß.

Krömker, U. (2019). §1 FGO. In O.-G. Lippross & W. Seibel (Hrsg.), *Basiskommentar Steuerrecht.* Dr. Otto Schmidt.

Leipold, C. (2019). §13 UStG. In *Sölch/Ringleb Umsatzsteuergesetz.* München, C.H. Beck.

Leonard, A. & Robisch, M. (2020). §3 UStG. In *Bunjes UStG.* München, C.H. Beck.

Lippross, O.-G. & Seibel, W. (Hrsg.). (2019). *Basiskommentar Steuerrecht* (Bd. 88. Ergänzungslieferung). Dr. Otto Schmidt.

Löwisch, G. (2020). §84 HGB. In C. T. e. a. Ebenroth (Hrsg.), *Handelsgestzbuch Band 1.* C.H. Beck/Verlag Vahlen.

Martin, S. (2020a). §3 UStG. In *Sölch/Ringleb Umsatzsteuergesetz* (Bd. 89. EL). München, C.H. Beck.

Martin, S. (2020b). §3 UStG. In *Sölch/Ringleb Umsatzsteuergesetz* (Bd. 90. EL). München, C.H. Beck.

Meißner, A. (2020). Praxisprobleme bei Mehrzweckgutscheinen in (grenzüberschreitenden) Vertriebsketten: Zugleich eine erste Auseinandersetzung mit dem BMF-Schreiben v. 2.11.2020. *Umsatzsteuer Rundschau*, 914–918.

Michel, G. (2018). Internet-Marktplätze und Gutscheine im Fokus der Steuergesetzgebung - umsatzsteuerrechtliche Änderungen zum 01.01.2019. *juris – Die Monatszeitschrift*, 431–437.

Nieskens, H. (2010). Verkauf von Warengutscheinen: -EuGH-Entscheidung in Sachen Astra Zeneca und ihre Folgen für das deutsche Recht. *Umsatzsteuer Rundschau*, 893–902.

Nieskens, H. (2019a). §13 UStG. In *Rau/Dürrwächter UStG* (Bd. 183. Lieferung). Köln, Dr. Otto Schmidt.

Nieskens, H. (2019b). §3 UStG. In *Rau/Dürrwächter UStG* (Bd. 189. Lieferung). Köln, Dr. Otto Schmidt.

Nieskens, H. (2019c). §3 UStG. In *Rau/Dürrwächter UStG* (Bd. 191. Lieferung). Köln, Dr. Otto Schmidt.

Prätzler, R. C. (2020). BMF nimmt Stellung zu umsatzsteuerlichen Fragen bei Gutscheinen. *DER BETRIEB* (47), 2491–2493.

Prätzler, R. C. (2021). Umsatzsteuer bei Gutscheinen ab 1.1.2019 - Bestandsaufnahme zu offenen Fragen: Zugleich Anmerkungen zu BMF-Schreiben v. 2.11.2020. *StuB* (3), 107–112.

Ranzinger, A. (2017). *Praxiswissen Kundenbindungsprogramme: Konzeption und operative Umsetzung* (2. Aufl.). Wiesbaden, Springer Gabler.

Ratschow, E. (2020). §42 AO. In *Klein AO.* München: C.H. Beck.

Rau, G., Dürrwächter, E., Flick, H. & Geist, R. (Hrsg.). (2019). *Rau/Dürrwächter UStG*. Köln, Dr. Otto Schmidt.

Reiß, W., Kraeusel, J. & Langer, M. (2019). *Reiß/Kraeusel/Langer UStG*. Bonn, Stollfuß.

Robisch, M. (2020). §1 UStG. In *Bunjes UStG.* München, C.H. Beck.

Robisch, M. & Greif, S. (2018). Entwurf eines "JStG 2018": Die umsatzsteuerlichen Neuerungen aus Sicht der Praxis: Auswirkung und offene Fragen. *NWB – Steuer- und Wirtschaftsrecht* (37), 2682–2694.

Rust, M. & Tillmanns, J. (2020). Zu der umsatzsteuerlichen Behandlung von Einzweck-Gutscheinen im Lichte der vorübergehenden Mehrwertsteuersenkung. *USt direkt digital* (21), 17–19.

Slapio, U. & Feil, K. (2020). Temporäre Senkung der Mehrwertsteuersätze - Gestaltungsmöglichkeiten durch Ausgabe von Einzweck-Gutscheinen? *Umsatzsteuer Rundschau*, 703–706.

Sölch, O. & Ringleb, K. (2020). *Sölch/Ringleb Umsatzsteuergesetz*. München, C.H. Beck.

Stadie, H. (2019). §17 UStG. In *Rau/Dürrwächter UStG.* Köln, Dr. Otto Schmidt.

Sterzinger, C. (2020). Umsatzsteuerliche Behandlung von touristischen Gästekarten-Modellen. *Der Umsatz-Steuer-Berater*, 94–100.

Strieder, T. & Gahlen, D. (2020). §336 HGB. In M. Henssler (Hrsg.), *beck-online.GROSSKOMMENTAR.* München, C.H. Beck.

Thiele, P. J. & König, I. (2018). Die umsatzsteuerrechtliche Behandlung von Gutscheinen: Eine kritische Betrachtung der Neuregelungen. *Umsatzsteuer Rundschau*, 933–940.

Tipke, K. & Kruse, H. W. (2021). *Tipke/Kruse, AO/FGO*. Köln, Verlag Dr. Otto Schmidt.

VAT Committee. (2019a). *Information Paper v. 3.4.2019: VAT treatment of "city cards".* https://circabc.europa.eu/ui/group/cb1eaff7-eedd-413d-ab88-94f761f9773b/library/20546cf6-42a6-4196-b14f-d875b3826c06/details. (Zuletzt abgerufen am: 28.04.2021)

VAT Committee. (2019b). *Working paper No 983 v. 13.11.2019: Questions raised following implementation of the Voucher Directive.* https://circabc.europa.eu/ui/group/cb1eaff7-eedd-413d-ab88-94f761f9773b/library/8ded663d-0886-4d3a-b036-2d8c77e5d66a/details. (Zuletzt abgerufen am: 28.04.2021)

Vobbe, R. (2018). Ausgewählte umsatzsteuerliche Aspekte des "JStG 2018". *Betriebs-Berater*, 2844–2851.

Vobbe, R. & Stelzer, T. (2020). Gutscheine als Liquiditätsquelle: Umsatzsteuerrechtliche Auswirkungen von COVID-19 auf die Reiseverkehrs- und Veranstaltungsbranche. *DER BETRIEB* (21), 1086–1092.

Wäger, C. (2020). §3a UStG. In *Sölch/Ringleb Umsatzsteuergesetz* (Bd. 90. EL). München, C.H. Beck.

Weimann, R. (2020). Umsatzsteuer: Befristete Absenkung der USt für Gastronomen - Checkliste, Fälle und Praxistipps. *Betriebswirtschaft im Blickpunkt*, 179–183.

Weymüller, R. (Hrsg.). (2019). *Beck'scher Online-Kommentar UStG (BeckOK UStG)*. München, C.H. Beck.

Junior Management Science 8(1) (2023) 219-236

Junior Management Science

journal homepage: www.jums.academy

Is Cash (the only) King? – A critical analysis of the relevance of Cashflow figures according to IAS 7

Is Cash (the only) King? – Eine kritische Analyse der Aussagekraft von Kapitalflussrechnungen nach IAS 7

Jonas Pütter

Heinrich-Heine-Universität Düsseldorf

Abstract

When accountants or academics state „Cash is King“, they usually want to emphasize the practical relevance of Cashflow figures for external addressees of annual financial statements. In particular, Cashflow figures are considered relatively free from accounting policy measures and therefore appear to be particularly suitable indicators for measuring corporate success. Within this article, the actual relevance of Cashflow figures according to IAS 7 is evaluated. For this purpose, the influence of accounting policy measures on Cashflow figures according to IAS 7 is assessed. In addition, literature is examined that directly compares the ability of Cashflow figures and earnings figures to represent the success of a company. An important result of this article is that companies reduce the informative value of their Cashflow statement, when they decide to use the indirect method for the Cashflow from operating activities. For this reason, an obligation to use the direct method, when presenting the Cashflow from operating activities, would be desirable.

Zusammenfassung

Die Aussage „Cash is King“ wird häufig getätigt, um eine besondere praktische Relevanz von Cashflow-Größen der Kapitalflussrechnung für den externen Adressaten des Jahresabschlusses hervorzuheben. Cashflow-Größen wird insbesondere eine geringe Beeinflussbarkeit durch Bilanzpolitik zugesprochen und somit eine besondere Eignung als Maß des Unternehmenserfolgs. Im Rahmen dieser Arbeit wird die tatsächliche Aussagekraft von Cashflow-Größen der Kapitalflussrechnung nach IAS 7 eingeordnet. Dazu wird zum einen die Beeinflussbarkeit der Kapitalflussrechnung durch bilanzpolitische Maßnahmen untersucht. Zum anderen wird Literatur betrachtet, welche die Aussagekraft von Cashflow- und Ergebnis-Größen über den Erfolg eines Unternehmens gegenüberstellt. Ein wichtiges Resultat der Arbeit ist, dass Unternehmen durch die Ausübung des Wahlrechtes zur indirekten Darstellung des Cashflows aus betrieblicher Tätigkeit, die Aussagekraft ihrer Kapitalflussrechnung beschränken. Aus diesem Grunde wäre eine Verpflichtung zur Veröffentlichung einer direkten Darstellung des Cashflows aus betrieblicher Tätigkeit wünschenswert.

Keywords: IAS 7; Cashflow; Kapitalflussrechnung; Bilanzpolitik.

1. Einleitung

Die Aussage „Cash is King“ impliziert eine besondere praktische Relevanz von Cashflow-Größen der Kapitalflussrechnung für den externen Adressaten des Jahresabschlusses. Ein zentrales Argument für die Betrachtung solcher Zahlungsströme ist, dass Zahlungsströme wenig anfällig gegenüber Bilanzpolitik durch das bilanzierende Unternehmen sind und somit insgesamt den Unternehmenserfolg adäquat abbilden (vgl. Gebhardt & Mansch, 2012, S. 130f.). Andererseits wird in der Fachliteratur betont, dass Cashflow-Größen nur bedingt als Erfolgsindikator betrachtet werden sollten, da betriebswirtschaftlich begründete Periodenabgrenzungen, wie zum Beispiel Abschreibungen, anders als im Periodenergebnis der Gesamtergebnisrechnung, in den Cashflow-Größen der Kapitalflussrechnung nicht berücksichtigt werden (vgl. Lachnit, 2004, S. 253). Hinsichtlich der

DOI: https://doi.org/10.5282/jums/v8i1pp219-236

Aussagekraft der Cashflow-Größen aus der Kapitalflussrechnung, insbesondere auch im Vergleich zur Aussagekraft des Periodenergebnisses, existieren somit abweichende Einschätzungen. Eine korrekte Einordnung der Aussagekraft der Kapitalflussrechnung nach International Accounting Standard (IAS) 7 (IAS, 2016) ist jedoch von hoher Bedeutung. So zeigen Hitz und Teuteberg (2013, S. 40) in einer empirischen Studie, dass etwa 50 Prozent der im Prime Standard an der Frankfurter Wertpapierbörse gelisteten Unternehmen, eine Cashflow-Größe der Kapitalflussrechnung nach IAS 7 im Rahmen ihrer freiwilligen Berichterstattung kommunizieren. Ein verpflichtender Bestandteil des Jahresabschlusses ist die Kapitalflussrechnung nach IAS 7 ohnehin bereits für eine Großzahl an Unternehmen, seit im Jahr 2005 die Bilanzierung nach den International Financial Reporting Standards (IFRS) für kapitalmarktorientierte Konzerne mit Sitz in Deutschland vorgeschrieben wurde (vgl. Lüdenbach, Hoffmann & Freiberg, 2020, § 7, Rn. 6). Vor diesem Hintergrund kann geschlossen werden, dass Cashflow-Informationen aus der Kapitalflussrechnung nach IAS 7 auch in Deutschland von Adressaten des Jahresabschlusses signifikant genutzt werden. Für welche Zwecke diese Nutzung in der Praxis erfolgt, wird als erste Forschungsfrage grundlegend in Kapitel 3 geklärt.

Im Verlauf der Arbeit soll die Aussagekraft der Kapitalflussrechnung nach IAS 7 weitergehend eingeordnet werden, indem die genannten, zentralen Thesen in Bezug auf die Aussagekraft der Kapitalflussrechnung überprüft werden: In Kapitel 4 wird untersucht, ob die Kapitalflussrechnung nach IAS 7 tatsächlich nur in geringem Maße bilanzpolitisch manipulierbar ist, was der Aussagekraft der Kapitalflussrechnung zugutekommen würde. Anschließend wird in Kapitel 5 überprüft, ob Periodenabgrenzungen die Aussagekraft des Periodenergebnisses gegenüber der Aussagekraft etwaiger Cashflow-Größen der Kapitalflussrechnung erhöhen, sodass der Unternehmenserfolg tatsächlich besser durch das Periodenergebnis abgebildet wird. Anhand dieser beiden weiterführenden Forschungsfragen wird die Aussagekraft der Kapitalflussrechnung nach IAS 7 im Rahmen dieser Arbeit also umfassend beleuchtet, indem die Aussagekraft der Kapitalflussrechnung sowohl vor dem Hintergrund bilanzpolitischer Maßnahmen, als auch im relativen Vergleich zum Periodenergebnis, als weitere zentrale Größe der Rechnungslegung, untersucht wird. Ein wichtiges Resultat der Arbeit ist, dass die direkte Darstellung des Cashflows aus betrieblicher Tätigkeit der indirekten Darstellung hinsichtlich ihrer Aussagekraft überlegen ist, weshalb eine Verpflichtung zur Veröffentlichung einer direkten Darstellung des Cashflows aus betrieblicher Tätigkeit wünschenswert wäre. Im folgenden Kapitel werden zunächst Struktur und Inhalt der Kapitalflussrechnung nach IAS 7 grundlegend erläutert.

2. Inhalt und Struktur der Kapitalflussrechnung nach IAS 7

2.1. Fondsveränderungsrechnung

Im Rahmen der Kapitalflussrechnung werden nicht zahlungswirksame Periodenabgrenzungen, die Bewertungsspielräumen unterliegen, im Vergleich zum Periodenergebnis nicht berücksichtigt, sodass ausschließlich Zahlungsströme (im Folgenden "Cashflows" genannt) betrachtet werden (vgl. Lüdenbach et al., 2020, § 3, Rn. 3). Die daraus resultierende Darstellung der Kapitalflussrechnung nach IAS 7 setzt sich aus einer Ursachenrechnung und einer Fondsveränderungsrechnung zusammen (vgl. Lüdenbach et al., 2020, § 3, Rn. 35). Die Fondsveränderungsrechnung stellt die Gesamtveränderung des Finanzmittelfonds innerhalb der betrachteten Periode dar. Diese Gesamtveränderung des Finanzmittelfonds ergibt sich aus der Summe der zahlungswirksamen Fondsveränderungen, die aus der Ursachenrechnung zu entnehmen ist, zuzüglich zahlungsunwirksamer Fondsveränderungen (vgl. Hitz & Teuteberg, 2013, S. 35). Letztere können sich beispielsweise durch Änderungen im Wechselkurs oder im Konsolidierungskreis ergeben (vgl. Eiselt & Müller, 2014, S. 50).

Der Finanzmittelfonds wird nach IAS 7 eng definiert und setzt sich ausschließlich zusammen aus Zahlungsmitteln und Zahlungsmitteläquivalenten (vgl. Lüdenbach et al., 2020, § 3, Rn. 15f.). Die Zahlungsmittel sind gemäß IAS 7.6 definiert als Summe der Barmittel und Sichteinlagen, während die Zahlungsmitteläquivalente sich zusammen setzten aus hochliquiden „Finanzinvestitionen, die jederzeit in festgelegte Zahlungsmittelbeträge umgewandelt werden können und nur unwesentlichen Werteschwankungsrisiken unterliegen" (IAS 7.6). Eine Finanzinvestition qualifiziert sich normalerweise nur dann für einen Einbezug in die Zahlungsmitteläquivalente, wenn sie ab dem Zeitpunkt des Erwerbs eine Restlaufzeit besitzt, die nicht mehr als drei Monate beträgt (vgl. IAS 7.7). Darüber hinaus dürfen Zahlungsmitteläquivalente in der Regel nur mit dem Ziel gehalten werden, kurzfristige Zahlungsverpflichtungen zu erfüllen. Das heißt Finanzmittel, die für andere Zwecke gehalten werden, zählen allgemein nicht zu den Zahlungsmitteläquivalenten (vgl. IAS 7.7). Grundsätzlich besteht der Finanzmittelfonds somit aus Posten, die auf der Aktivseite der Bilanz ausgewiesen werden. Nach IAS 7.8 werden jedoch unter bestimmten Umständen auch Kontokorrentkredite dem Finanzmittelfonds zugerechnet, wodurch sich die Höhe des Finanzmittelfonds entsprechend reduziert. Aufgrund der generalisierenden Formulierungsweise sowohl im IAS 7.7 als auch im IAS 7.8, werden dem Unternehmen hier gewisse Ermessensspielräume bei der Abgrenzung des Finanzmittelfonds gewährt (vgl. Lüdenbach et al., 2020, § 3, Rn. 19). Inwiefern diese Spielräume möglicherweise die Aussagekraft der Kapitalflussrechnung beschränken, wird in Kapitel 4.3 untersucht.

2.2. Ursachenrechnung

Die Ursachenrechnung ist der zentrale Bestandteil der Kapitalflussrechnung. Im Rahmen der Ursachenrechnung wird

die zahlungswirksame Veränderung von Posten dargestellt, die nicht zum Finanzmittelfonds hinzu zählen (vgl. Lüdenbach et al., 2020, § 3, Rn. 35). Daraus wird für den externen Adressaten des Jahresabschlusses ersichtlich, aus welchen Quellen eine Erhöhung oder Minderung des Finanzmittelfonds resultiert und darüber hinaus auch, für welche Zwecke die generierten Mittel verwendet werden (vgl. Eiselt & Müller, 2014, S. 37). Die Mittelherkunft und die Mittelverwendung werden im Rahmen der Ursachenrechnung in drei Teilbereichen dargestellt: Unterschieden wird zwischen dem Cashflow aus betrieblicher Tätigkeit, dem Cashflow aus Investitionstätigkeit sowie dem Cashflow aus Finanzierungstätigkeit (vgl. IAS 7.10).

Der Cashflow aus betrieblicher Tätigkeit setzt sich grundsätzlich aus den operativen Ein- und Auszahlungen des Unternehmens zusammen (vgl. Ruprecht & Nösberger, 2012, S. 109). Dazu zählen gemäß IAS 7.14 beispielsweise Einzahlungen aus Güterverkauf und erbrachten Dienstleistungen oder Auszahlungen an Lieferanten und Mitarbeiter. Die Cashflow-Größe ermöglicht es dem Adressaten einzuschätzen, ob die laufenden Auszahlungen des operativen Geschäftes durch die laufenden Einzahlungen gedeckt werden können und inwiefern etwaige Überschüsse für die Investitions- und Finanzierungstätigkeit zur Verfügung stehen (vgl. Ordelheide & Leuz, 1998, S. 182). Die Deckung der operativen Auszahlungen wird von Seite des Adressaten häufig als Mindestkriterium definiert, wenn es um die Beurteilung der Finanzkraft des laufenden Geschäftes des Unternehmens geht (vgl. Rehkugler & Poddig, 1998, S. 129). Bei der Beurteilung zu berücksichtigen sind jedoch auch Spezifika des Unternehmens, so schaffen es Unternehmen in der Einführungsphase typischerweise nicht, ihre operativen Auszahlungen eigenständig zu decken (vgl. Oehlrich, 2020, S. 356).

Hinsichtlich der Darstellung des Cashflows aus betrieblicher Tätigkeit besteht ein explizites Wahlrecht (vgl. IAS 7.18) zwischen der direkten Darstellung und der indirekten Darstellung, wobei eine Empfehlung für die direkte Darstellung ausgesprochen wird (vgl. IAS 7.19). Die direkte Darstellung beginnt typischerweise mit den Umsatzeinzahlungen und führt dann weitere Aufwendungen und Erträge auf, die bei Zustandekommen einen Cashflow bewirkt haben (vgl. Lachnit & Müller, 2017, S. 299). Im Rahmen der indirekten Darstellung hingegen wird das Periodenergebnis, durch eine Bereinigung um nicht zahlungswirksame Aufwendungen und Erträge, Cashflows aus der Investitions- oder Finanzierungstätigkeit und Bestandsveränderungen des Nettoumlaufvermögens, in den Cashflow aus betrieblicher Tätigkeit überführt (vgl. IAS 7.20). Darüber hinaus lässt der IAS 7.20 Spielraum bei der Wahl der nicht näher definierten Ausgangsgröße "Periodenergebnis". Denkbar wäre somit beispielsweise ein Ergebnis nach Steuern, aber genauso auch ein Ergebnis vor Steuern, oder sogar ein Ergebnis vor Steuern, Zinsen und Abschreibungen (vgl. Hüttche & Int-Veen, 2008, S. 351).

Ihre Empfehlung zur Nutzung der direkten Darstellung des Cashflows aus betrieblicher Tätigkeit begründen die IFRS damit, dass die direkte Darstellung Informationen bereitstelle, die aus der indirekten Darstellung nicht entnommen werden können (vgl. IAS 7.19). Auch in der Literatur wird der direkten Darstellung eine höhere Aussagekraft zugesprochen, insbesondere da aus der direkten Darstellung die operativen Ein- und Auszahlungen durch den Adressaten direkt entnommen werden können (vgl. Meyer, 2007, S. 388). Andere bedeutende Wahlrechte im Rahmen der Kapitalflussrechnung betreffen die Zuordnung von Zins-, Dividenden- (vgl. IAS 7.33f.) sowie Ertragssteuerzahlungen (vgl. IAS 7.35f.), die grundsätzlich dem Cashflow aus betrieblicher Tätigkeit, aber auch den beiden anderen Cashflow-Größen der Ursachenrechnung zurechenbar sind. Ob und in welchem Maße in der Praxis das Wahlrecht zur indirekten Darstellung des Cashflows aus betrieblicher Tätigkeit sowie die genannten Zuordnungswahlrechte tatsächlich genutzt werden, wird in Kapitel 4.3 überprüft.

Der Cashflow aus Investitionstätigkeit, als zweite Cashflow-Größe der Ursachenrechnung, setzt sich zusammen aus Ein- und Auszahlungen für Veräußerung und Erwerb von langfristigen Vermögenswerten und Finanzinvestitionen (vgl. IAS 7.6). Da die Investitionen typischerweise höher ausfallen als die Desinvestitionen, ist der Cashflow in der Regel negativ und muss durch die Cashflows aus betrieblicher Tätigkeit oder Finanzierungstätigkeit gedeckt werden, damit der Finanzmittelfonds nicht reduziert wird (vgl. Rehkugler & Poddig, 1998, S. 130).

Als letzter Bestandteil der Ursachenrechnung umfasst der Cashflow aus Finanzierungstätigkeit Ein- und Auszahlungen in Bezug auf Fremd- und Eigenkapital (vgl. Lüdenbach et al., 2020, § 3, Rn. 89). Dazu gehören gemäß IAS 7.17 zum Beispiel die Einzahlungen aus Kapitalerhöhung oder aber die Auszahlungen für den Rückkauf von Unternehmensanteilen. Die Aufgabe des Cashflows aus Finanzierungstätigkeit ist es, eine potentiell auftretende Finanzierungslücke, die sich aus der Summe der Cashflows aus betrieblicher Tätigkeit und Investitionstätigkeit ergibt, zu füllen, sodass für die Finanzierung von operativem Geschäft und Investitionstätigkeit keine Mittel aus dem Finanzmittelfonds entnommen werden müssen (vgl. Rehkugler & Poddig, 1998, S. 130). Anders als für den Cashflow aus betrieblicher Tätigkeit wird für die Cashflows aus Investitionstätigkeit und Finanzierungstätigkeit die Darstellung in direkter Form vorgeschrieben (vgl. IAS 7.21).

Insgesamt liefert die Ursachenrechnung die zahlungswirksame Gesamtveränderung des Finanzmittelfonds in der betrachteten Periode als Summe der drei, nun erläuterten, Cashflow-Größen. Im folgenden Kapitel wird untersucht, wie auf Basis dieser drei Cashflow-Größen der Ursachenrechnung Kennzahlen gebildet werden können, mit denen die Finanz- und Ertragskraft eines Unternehmens analysiert, sowie der Wert eines Unternehmens ermittelt werden kann.

3. Externe Nutzung der Kapitalflussrechnung nach IAS 7 für Bilanzanalyse und Unternehmensbewertung

3.1. Kapitalflussrechnung als Indikator der Finanzkraft

Zweck der Bilanzanalyse ist es die gegenwärtige und zukünftige, ökonomische Zielerreichung des Unternehmens

zu beurteilen (vgl. Coenenberg, Haller & Schultze, 2021, S. 1085). Dabei liegt der Fokus, bei einer engen Definition der Bilanzanalyse, auf einer Beurteilung der finanzwirtschaftlichen und erfolgswirtschaftlichen Zielerreichung (vgl. Coenenberg et al., 2021, S. 1087), sodass die Analyse von Finanz- und Ertragskraft als zentraler Bestandteil der Bilanzanalyse konkretisiert werden kann. Gemäß den IFRS soll die Kapitalflussrechnung indes in erster Linie die folgenden drei Zwecke erfüllen: Sie soll als Indikator der Finanzkraft sowie als ergänzender Indikator der Ertragskraft dienen und darüber hinaus die Modellierung des Barwertes zukünftiger Cashflows erleichtern (vgl. IAS 7.4). Barwerte zukünftiger Cashflows sind in der Praxis häufig Inputfaktoren für die Unternehmensbewertung, wie im späteren Verlauf dieses Kapitels näher erläutert wird. Da die Kapitalflussrechnung nach IAS 7 also insbesondere eine Analyse der Finanz- und Ertragskraft ermöglichen soll, ist sie neben anderen Bestandteilen des Jahresabschlusses, wie Bilanz oder Gesamtergebnisrechnung, ein zentrales Instrument der Bilanzanalyse. Im Rahmen dieses Kapitels soll zuerst näher ausgeführt werden, wie die Cashflow-Größen der Kapitalflussrechnung in der Praxis konkret für die Analyse der Finanz- und Ertragskraft, und somit die Bilanzanalyse, genutzt werden können. Anschließend wird im dritten Teil des Kapitels die Nutzung der Kapitalflussrechnung als Grundlage für die Unternehmensbewertung konkretisiert.

In der Literatur wird der Kapitalflussrechnung in erster Linie eine wesentliche Bedeutung als Finanzindikator zugesprochen: Demnach ist eine umfangreiche Einschätzung der Liquiditätslage eines Unternehmens einzig anhand der Kapitalflussrechnung möglich (vgl. Pollmann, 2012, S. 191f.). Durch eine Analyse der Finanzkraft des Unternehmens soll der Adressat beurteilen können, ob das Unternehmen gegenwärtig und zukünftig in der Lage ist, seinen finanziellen Verpflichtungen nachzukommen. Dazu gehört insbesondere die Sicherstellung dauerhafter Zahlungsfähigkeit (vgl. Lachnit & Müller, 2017, S. 5). Wichtige Warnsignale, die im Rahmen der Kapitalflussrechnung auf eine geringe Finanzkraft hindeuten, sind neben einem defizitären Cashflow aus betrieblicher Tätigkeit auch eine geringe Investitionstätigkeit, sowie eine starke Außenfinanzierung mittels Fremdkapital (vgl. Rehkugler & Poddig, 1998, S. 131).

Grundsätzlich kann für die Analyse der Finanzkraft eine absolute Kennzahl, wie der Cashflow aus betrieblicher Tätigkeit, betrachtet werden. Von größerer Bedeutung für den Zweck der Bilanzanalyse sind jedoch relative Kennzahlen, da erst die unternehmensspezifische, relative Betrachtung einen Schluss darauf zulässt, ob die absolute Cashflow-Größe eine angemessene Höhe aufweist oder nicht (vgl. Meyer, 2007, S. 521). Im Folgenden soll konkreter dargestellt werden, welche relativen Kennzahlen gebildet werden können, um verschiedene Aspekte der Finanzkraft zu analysieren. Der Rückgriff auf Kennzahlen ist dabei ein typisches Mittel im Rahmen der Bilanzanalyse, da Kennzahlen den Vorteil bieten, die betrachteten Sachverhalte prägnant und intuitiv darzustellen (vgl. Eiselt & Müller, 2014, S. 117). Die wichtigste Cashflow-Größe für die Analyse der Finanzkraft ist der Cashflow aus betrieblicher Tätigkeit: Durch Analyse des Cashflows aus betrieblicher Tätigkeit kann der Adressat erkennen, ob das Unternehmen in der Lage war einen Einzahlungsüberschuss aus dem operativen Geschäft zu erwirtschaften um damit, ohne externe Mittel oder Desinvestition, essentiellen Aufgaben wie Investition, Schuldentilgung und Gewinnausschüttung nachzukommen (vgl. Müller, 2008, S. 293). Der Cashflow aus betrieblicher Tätigkeit befähigt den Adressaten somit insbesondere zu einer Analyse der Innenfinanzierungskraft, Schuldentilgungskraft und Gewinnausschüttungskraft.

Die nachhaltige Innenfinanzierungskraft eines Unternehmens umschreibt die Fähigkeit aus operativen Einzahlungen liquide Mittel für die Investitionstätigkeit zu erwirtschaften (vgl. Meyer, 2007, S. 524f.). Zur Bemessung der Innenfinanzierungskraft kann die in Abbildung (Abb.) 1 dargestellte Investitionsdeckung betrachtet werden.

Die Kennzahl lässt sich berechnen als Quotient aus Cashflow aus betrieblicher Tätigkeit und Cashflow aus Investitionstätigkeit. Da der Cashflow aus Investitionstätigkeit in der Regel negativ ausfällt, sollte dieser Cashflow bei der Quotientenbildung entsprechend mit -1 multipliziert werden (vgl. Kaiser & Bierwulf, 2011, S. 11). Ist die Investitionsdeckung größer oder gleich eins kann der Adressat schließen, dass dem Unternehmen ausreichend operative Einzahlungen zugeflossen sind, um die Auszahlungen für die Investitionstätigkeit zu decken.

Nach Analyse der Schuldentilgungskraft sollte der Adressat indes einschätzen können, ob und wie schnell das Unternehmen seine Finanzschulden aus dem eigenen, operativen Geschäft heraus decken kann (vgl. Meyer, 2007, S. 537). Eine Kennzahl die für diesen Zweck betrachtet wird ist der dynamische Verschuldungsgrad, der gegeben ist durch den Quotienten aus Finanzschulden und Cashflow aus betrieblicher Tätigkeit (vgl. Abb. 2). Je nach Analyseziel des Adressaten können dabei ausschließlich langfristige Finanzschulden einbezogen, oder Zahlungsmittelbestände wie der Kassenbestand von den Finanzschulden abgezogen werden (vgl. Meyer, 2007, S. 539), was den Verschuldungsgrad und damit die Tilgungsdauer entsprechend reduzieren würde.

Zuletzt beschreibt die Gewinnausschüttungskraft die Fähigkeit eines Unternehmens, seine Gewinnausschüttungen aus dem operativen Geschäft zu bestreiten. Als zugehörige Kennzahl kann insbesondere die in Abb. 3 dargestellte Dividendendeckung betrachtet werden, die sich ergibt bei Division des Cashflows aus betrieblicher Tätigkeit durch die gezahlten Dividenden (vgl. Meyer, 2007, S. 546).

Sofern gezahlte Zinsen dem Cashflow aus betrieblicher Tätigkeit, gemäß dem Wahlrecht nach IAS 7.33, nicht bereits zugeordnet wurden, sollten sie bei der Quotientenbildung nachträglich vom Cashflow subtrahiert werden, da Zinszahlungen durch das Unternehmen in jedem Fall zu leisten sind und Vorrang haben vor der Dividendenauszahlung (vgl. Meyer, 2007, S. 546). Erstrebenswert ist eine mehrfache Deckung der Dividendenzahlungen durch die operativen Einzahlungen. Eine Dividendendeckung kleiner eins zeigt indes an, dass die geleisteten Dividendenzahlungen nicht über das operative Geschäft gedeckt werden können (vgl. Meyer,

$$\text{Investitionsdeckung} = \frac{\text{Cashflow aus betrieblicher Tätigkeit}}{-1 * (\text{Cashflow aus Investitionstätigkeit})}$$

Abbildung 1: Kennzahl Investitionsdeckung

vgl. Kaiser & Bierwulf, 2011, S.11

$$\text{Dynamischer Verschuldungsgrad} = \frac{\text{Finanzschulden}}{\text{Cashflow aus betrieblicher Tätigkeit}}$$

Abbildung 2: Kennzahl Dynamischer Verschuldungsgrad

vgl. Meyer, 2007, S. 537

$$\text{Dividendendeckung} = \frac{\text{Cashflow aus betrieblicher Tätigkeit} - \text{Zinsauszahlungen}}{\text{Dividendenauszahlungen}}$$

Abbildung 3: Kennzahl Dividendendeckung

vgl. Meyer, 2007, S. 546

$$\text{Außenfinanzierungsgrad} = \frac{\text{Cashflow aus betrieblicher Tätigkeit}}{\text{Cashflow aus Finanzierungstätigkeit}}$$

Abbildung 4: Kennzahl Außenfinanzierungsgrad

vgl. Meyer, 2007, S. 531

2007, S. 546).

Unter weitergehender Betrachtung der Cashflows aus Investitionstätigkeit und Finanzierungstätigkeit lässt sich darüber hinaus die Fähigkeit des Unternehmens beurteilen, liquide Mittel außerhalb des operativen Geschäftes, durch Desinvestition und Außenfinanzierung, zu erwirtschaften (vgl. Meyer, 2007, S. 519). Von besonderer Relevanz ist dabei die Analyse der Außenfinanzierungskraft: Die Außenfinanzierung umfasst die gesamte unternehmensexterne Finanzierungstätigkeit mittels Eigen- oder Fremdkapitalaufnahme (vgl. Meyer, 2007, S. 530). Als Kennzahl für die Beurteilung der Außenfinanzierungskraft kann der Außenfinanzierungsgrad (vgl. Abb. 4) betrachtet werden, der gegeben ist durch den Quotienten aus Cashflow aus betrieblicher Tätigkeit und Cashflow aus Finanzierungstätigkeit (vgl. Meyer, 2007, S. 531).

Die Interpretation dieser Kennzahl ist weniger intuitiv als bei den bisher vorgestellten Kennzahlen, da die Kennzahl bei unterschiedlichen Vorzeichen der Cashflow-Größen unterschiedliche Rückschlüsse zulässt (vgl. Meyer, 2007, S. 531). Für den Adressaten ist jedoch insbesondere von Interesse, inwiefern ein Finanzierungsdefizit im operativen Bereich durch Außenfinanzierung gedeckt werden kann: Im Szenario eines negativen Cashflows aus betrieblicher Tätigkeit und positiven Cashflows aus Finanzierungstätigkeit gibt die Kennzahl Auskunft darüber, ob die Außenfinanzierungstätigkeit ausreichend war, um ohne Rückgriff auf den Finanzmittelfonds das Defizit an operativen Einzahlungen zu überbrücken. Damit dies der Fall ist, muss der Absolutbetrag der Kennzahl kleiner oder gleich eins sein. In dem Falle lässt sich darüber hinaus eine Aussage über den Anteil der aufgenommenen Finanzmittel treffen, der zum Ausgleich des Einzahlungsdefizites aus dem operativen Bereich eingesetzt werden musste (vgl. Meyer, 2007, S. 531).

Die Spezialisierung der nun vorgestellten Kennzahlen auf einen Aspekt der Finanzkraft befähigt den Adressaten, spezifische Fragestellungen zielgerichtet zu beantworten. Gleichzeitig ist ein zusammenfassendes Urteil über die Finanzkraft des Unternehmens nur anhand einer Kennzahl möglich, die den gesamten Finanzierungsbedarf einer Periode berücksich-

tigt. Genau das leistet die “Cashflow Adequacy Ratio” (vgl. Meyer, 2007, S. 548), die in der untenstehenden Abb. 5 dargestellt wird.

Sie gibt an, ob mit dem Cashflow aus betrieblicher Tätigkeit der gesamte Finanzierungsbedarf der betrachteten Periode gedeckt werden konnte, was durch einen Wert größer oder gleich eins zum Ausdruck gebracht wird (vgl. Meyer, 2007, S. 548). Ein Wert kleiner als eins signalisiert hingegen ein Finanzierungsdefizit, das durch Außenfinanzierung oder durch Rückgriff auf Mittel aus dem Finanzmittelfonds geschlossen werden muss (vgl. Meyer, 2007, S. 548).

3.2. Kapitalflussrechnung als Indikator der Ertragskraft

Durch eine Analyse der Ertragskraft soll der Adressat beurteilen können, ob das Unternehmen gegenwärtig in der Lage ist und zukünftig in der Lage sein wird, Erfolge zu erwirtschaften, um dadurch seine wirtschaftliche Leistungsfähigkeit zu erhalten oder zu verbessern (vgl. Lachnit & Müller, 2017, S. 5). Wie im vorangegangenen Abschnitt ausgeführt, wird in der Literatur insbesondere die Bedeutung der Kapitalflussrechnung als Indikator der Finanzkraft gewürdigt. Für die Beurteilung der Ertragskraft wird hingegen häufig in erster Linie das Periodenergebnis der Gesamtergebnisrechnung betrachtet: Demnach ist die Kapitalflussrechnung nur bedingt als Erfolgsindikator heran zu ziehen, weil betriebswirtschaftlich begründete Periodenabgrenzungen, wie zum Beispiel Abschreibungen, in den Cashflow-Größen nicht berücksichtigt werden (vgl. Wulf, 2008, S. 649). Da die Cashflow-Größen der Kapitalflussrechnung gleichzeitig als nur geringfügig bilanzpolitisch manipulierbar wahrgenommen werden (vgl. Meyer, 2007, S. 558f.), kann die Kapitalflussrechnung als nachrangiges Analyseinstrument zur Plausibilisierung des Periodenergebnissen dienen. Der Kapitalflussrechnung kommt demnach im Rahmen der erfolgswirtschaftlichen Analyse eine, im Vergleich zum Periodenergebnis, sekundäre Bedeutung zu. Entgegen dieser Sichtweise werden in der Praxis verstärkt auch Cashflow-Größen als primäre Indikatoren des Unternehmenserfolges betrachtet, da Szenarien existieren, in denen der Cashflow den Unternehmenserfolg möglicherweise sogar besser abbildet, als das Periodenergebnis (vgl. DeFond & Hung, 2003, S. 88). Aus diesen unterschiedlichen Sichtweisen ergibt sich die Notwendigkeit einer Analyse, ob das Periodenergebnis, aufgrund der Berücksichtigung von Periodenabgrenzungen, tatsächlich besser für die Abbildung des Unternehmenserfolges geeignet ist und inwiefern Szenarien existieren, in denen Cashflow-Größen für sich genommen eine höhere Aussagekraft aufweisen. Diese Analyse erfolgt in Kapitel 5. Im weiteren Verlauf dieses Abschnittes wird in erster Linie die Plausibilisierungsfunktion der Kapitalflussrechnung betrachtet.

Zur Plausibilisierung des Periodenergebnisses werden spezifische, relative Kennzahlen auf Grundlage der Kapitalflussrechnung gebildet, anhand derer der Adressat beurteilen kann, ob das Periodenergebnis als primärer Indikator der Ertragskraft durch signifikante, bilanzpolitische Maßnahmen beeinflusst wurde. Eine Kennzahl die für diesen Zweck betrachtet werden kann, ist die Umsatzqualität (vgl. Abb. 6). Sie ist gegeben durch den Quotienten aus Umsatzeinzahlungen und Umsatzerlös und lässt somit ein Urteil darüber zu, welcher Anteil der in der Gesamtergebnisrechnung ausgewiesenen Umsatzerlöse, dem Unternehmen in der betrachteten Periode tatsächlich durch Einzahlungen zugeflossen ist (vgl. Meyer, 2007, S. 582f.).

Erstrebenswert ist ein möglichst hoher Wert nahe eins, da in dem Falle davon ausgegangen werden kann, dass bilanzpolitische Spielräume, welche etwa im Falle von Langfristfertigungen vorliegen, die in der Gesamtergebnisrechnung ausgewiesenen Umsatzerlöse nicht signifikant verzerren (vgl. Meyer, 2007, S. 583). Wichtig anzumerken ist jedoch, dass die Umsatzeinzahlungen, als Komponente des Cashflows aus betrieblicher Tätigkeit, nur bei der direkten Darstellung dieses Cashflows aus der Kapitalflussrechnung unmittelbar entnommen werden können (vgl. Meyer, 2007, S. 582).

Eine weitere Möglichkeit das Periodenergebnis zu plausibilisieren, besteht in der Bildung von Rentabilitätskennzahlen. Rentabilitätskennzahlen werden in erster Linie bei der erfolgswirtschaftlichen Analyse mittels Gesamtergebnisrechnung gebildet, indem das Periodenergebnis im Verhältnis zum eingesetzten Kapital betrachtet wird. Auf Basis der Kapitalflussrechnung können ähnliche Rentabilitätskennzahlen konstruiert werden, die eine Cashflow-Größe im Verhältnis zum eingesetzten Kapital betrachten (vgl. Eiselt & Müller, 2014, S. 118). Durch Vergleich der Cashflow-basierten Kennzahl mit der Periodenergebnis-basierten Kennzahl, lässt sich eine Einschätzung über den Einfluss bilanzpolitischer Maßnahmen auf das Periodenergebnis treffen. Um diesem Zweck nachzukommen, muss die Cashflow-Größe im Zähler der Kennzahl, konzeptionell dem Periodenergebnis entsprechen. Exemplarisch kann die in der unten aufgeführten Abb. 7 dargestellte Cashflow-Eigenkapitalrentabilität betrachtet werden.

In den Nenner dieser Kennzahl wird, analog zur Periodenergebnis-basierten Kennzahl, das Eigenkapital gesetzt, während im Zähler der Cashflow aus betrieblicher Tätigkeit nach Cashflow aus Investitionstätigkeit, Zinsauszahlungen und gezahlten Ertragssteuern aufgeführt wird (vgl. Eiselt & Müller, 2014, S. 119). Festzuhalten ist, dass die Cashflow-Eigenkapitalrentabilität nicht exakt der Periodenergebnis-basierten Eigenkapitalrentabilität entsprechen kann. Durch vergleichende Betrachtung der Entwicklung dieser beiden Kennzahlen lässt sich allerdings einschätzen, ob die Periodenergebnis-basierte Eigenkapitalrentabilität dem Niveau der tatsächlichen Rentabilität des Eigenkapitals entspricht, oder ob signifikante Abweichungen vorliegen, die auf die Nutzung von Bewertungsspielräumen hindeuten (vgl. Meyer, 2007, S. 588).

Die Prämisse, die der Bedeutung der nun vorgestellten Kennzahlen als Plausibilisierungsinstrument zu Grunde liegt ist, dass die Cashflow-Größen der Kapitalflussrechnung selber kaum bilanzpolitisch manipulierbar sind. Ob diese Annahme gerechtfertigt ist, wird in Kapitel 4 überprüft.

$$\text{Cashflow Adequacy Ratio} = \frac{\text{Cashflow aus betrieblicher Tätigkeit}}{-1 * \left[\begin{array}{c}\text{Cashflow aus Investitionstätigkeit} \\ + \text{Auszahlungen zur Tilgung von Finanzschulden} \\ + \text{Zinsauszahlungen} + \text{Dividendenauszahlungen}\end{array}\right]}$$

Abbildung 5: Kennzahl Cashflow Adequacy Ratio

vgl. Meyer, 2007, S. 548

$$\text{Umsatzqualität} = \frac{\text{Umsatzeinzahlungen}}{\text{Umsatzerlöse}}$$

Abbildung 6: Kennzahl Umsatzqualität

vgl. Meyer, 2007, S. 583

$$\text{Cashflow-Eigenkapitalrentabilität} = \frac{\begin{array}{c}\text{Cashflow aus betrieblicher Tätigkeit} \\ + \text{Cashflow aus Investitionstätigkeit} \\ - \text{Zinsauszahlungen} - \text{Ertragssteuerzahlungen}\end{array}}{\text{Eigenkapital}}$$

Abbildung 7: Kennzahl Cashflow-Eigenkapitalrentabilität

vgl. Eiselt & Müller, 2014, S. 119

3.3. Kapitalflussrechnung als Grundlage für die Unternehmensbewertung

Als dritter wichtiger Zweck der Kapitalflussrechnung nach IAS 7, soll in diesem Kapitel die Nutzung der Kapitalflussrechnung für die Modellierung zukünftiger Cashflows untersucht werden. Die Modellierung zukünftiger Cashflows ist insbesondere für die Unternehmensbewertung mittels Discounted-Cashflow-Verfahren (DCF-Verfahren) notwendig (vgl. Oehlrich, 2020, S. 357). Im Rahmen der DCF-Verfahren, wird der Wert des Unternehmens durch Diskontierung zukünftiger Cashflow-Größen mit einem unternehmensspezifischen Kapitalkostensatz ermittelt (vgl. Nowak, 2003, S. 24f.). Grundsätzlich werden die DCF-Verfahren in Bruttoverfahren und Nettoverfahren unterschieden (vgl. Ballwieser & Hachmeister, 2016, S. 137f.). Im Rahmen des Bruttoverfahrens wird der Wert des Gesamtunternehmens ermittelt, das heißt es wird sowohl das Fremdkapital als auch das Eigenkapital bei der Wertermittlung berücksichtigt. Im Anschluss kann vom Gesamtunternehmenswert der Wert des Fremdkapitals abgezogen werden, um den Wert des Eigenkapitals zu erhalten (vgl. Schultze & Meyer, 2005, S. 293). Der Wert des Unternehmens aus Sicht aller Kapitalgeber setzt sich dabei zusammen, aus allen direkten und indirekten Cashflows, die den Kapitalgebern zufließen (vgl. Schultze & Meyer, 2005, S. 293). Während zu den direkten Cashflows zum Beispiel Gewinnausschüttungen oder Zinsauszahlungen gehören, entstehen indirekte Cashflows, wenn das Unternehmen überschüssige Zahlungsmittel einbehält und reinvestiert, sodass den Kapitalgebern zu einem späteren Zeitpunkt erhöhte Cashflows zukommen (vgl. Schultze & Meyer, 2005, S. 293). Als Zahlungsmittelbetrag, der potentiell zur Deckung dieser direkten und indirekten Cashflows zur Verfügung steht, kann der Brutto-Free-Cashflow (Brutto-FCF) betrachtet werden (vgl. Schultze & Meyer, 2005, S. 298f.). Aus der Kapitalflussrechnung nach IAS 7 kann der Brutto-FCF, für die Ermittlung des Gesamtunternehmenswertes, als Summe der Cashflows aus betrieblicher Tätigkeit und Investitionstätigkeit gebildet werden (vgl. Schultze & Meyer, 2005, S. 298f.). Da gemäß Kapitel 2.2 Zuordnungswahlrechte bezüglich Zins-, Dividenden- und Ertragssteuerzahlungen im Rahmen der Ursachenrechnung existieren, ist hier kritisch zu prüfen, wie die entsprechenden Wahlrechte ausgeübt wurden: Zins- und Dividendenauszahlungen dürfen nicht bereits in der Brutto-FCF-Größe erfasst sein und müssen gegebenenfalls heraus gerechnet werden (vgl. Schultze & Meyer, 2005, S. 313f.). Außerdem sollten Ertragssteuerzahlungen,

falls nicht bereits erfolgt, dem Cashflow aus betrieblicher Tätigkeit zugeordnet und somit in der Cashflow-Größe erfasst werden (vgl. Schultze & Meyer, 2005, S. 313f.). Durch Diskontierung zukünftiger Brutto-FCF kann der Adressat den Gesamtunternehmenswert bestimmen.

Beim Nettoverfahren hingegen wird der Wert des Eigenkapitals auf direkte Weise ermittelt, indem ausschließlich Cashflows betrachtet und diskontiert werden, die den Eigenkapitalgebern zufließen (vgl. Schultze & Meyer, 2005, S. 293). Nicht berücksichtigt werden somit fremdkapitalbezogene Cashflows. Für das Nettoverfahren wird der in Abb. 8 dargestellte Netto-Free-Cashflow (Netto-FCF) als Zahlungsmittelbetrag betrachtet, der sich entsprechend ergibt aus Brutto-FCF nach Veränderungen des Fremdkapitals und nach Zinsauszahlungen (vgl. Schultze & Meyer, 2005, S. 298f.).

Dabei müssen erneut, falls nicht bereits geschehen, Ertragssteuerzahlungen dem Cashflow aus betrieblicher Tätigkeit zugeordnet werden und Dividendenauszahlungen, sofern im Cashflow aus betrieblicher Tätigkeit enthalten, aus diesem heraus gerechnet werden (vgl. Schultze & Meyer, 2005, S. 313f.). Zinsauszahlungen werden, sofern sie nicht als Teil der Cashflows aus betrieblicher Tätigkeit oder Investitionstätigkeit erfasst wurden, als separate Abzugsgröße, entsprechend der in Abb. 8 definierten Formel verrechnet.

Trotz der erwähnten Zuordnungswahlrechte ist die Kapitalflussrechnung nach IAS 7 als Grundlage für die Ermittlung der beiden nun vorgestellten Free-Cashflow-Größen geeignet: Die Zins- (vgl. Lüdenbach et al., 2020, § 3, Rn. 109) und Dividendenzahlungen (vgl. Lüdenbach et al., 2020, § 3, Rn. 115) sowie die Ertragssteuerzahlungen (vgl. Lüdenbach et al., 2020, § 3, Rn. 120) müssen gesondert von anderen Cashflows im jeweiligen Cashflow-Bereich der Ursachenrechnung oder zumindest im Anhang des Jahresabschlusses angegeben werden, sodass der Adressat zur Ermittlung einer verzerrungsfreien Free-Cashflow-Größe, die betreffenden Cashflows leicht innerhalb der Kapitalflussrechnung umordnen kann. Somit ist eine sachgerechte Zuordnung der betreffenden Cashflows gewährleistet (vgl. Schultze & Meyer, 2005, S. 289f.).

Zusammenfassend lässt sich festhalten, dass die Kapitalflussrechnung nicht nur von großer Bedeutung für die Analyse der Finanzkraft ist, sondern auch als Grundlage für die Modellierung zukünftiger Cashflow-Größen im Rahmen der Unternehmensbewertung eine wichtige Rolle einnimmt. Im Rahmen der Analyse der Ertragskraft kommt ihr darüber hinaus, anhand der im vorherigen Abschnitt dargestellten Kennzahlen, eine Plausibilisierungsfunktion zu. Die Plausibilisierungsfunktion fußt dabei jedoch auf der Prämisse, dass die Kapitalflussrechnung selber kaum bilanzpolitisch manipulierbar ist. Diese Annahme muss, angesichts der in Kapitel 2 aufgeführten Wahlrechte und Spielräume im Rahmen des IAS 7, kritisch hinterfragt werden. Im folgenden Kapitel wird die Bedeutung dieser bilanzpolitischen Wahlrechte und Spielräume für die Aussagekraft der Kapitalflussrechnung nach IAS 7 näher untersucht.

4. Bedeutung von bilanzpolitischen Spielräumen für die Aussagekraft der Kapitalflussrechnung nach IAS 7

4.1. Einordnung der Bilanzpolitik in den Kontext der Kapitalflussrechnung

Die Kapitalflussrechnung nach IAS 7 wird, wie in Kapitel 3.2 erwähnt, grundsätzlich als ein Bestandteil des Jahresabschlusses wahrgenommen, der nur in geringem Maße bilanzpolitisch manipulierbar ist. Ein zentrales Argument das im Zusammenhang mit dieser These häufig genannt wird ist, dass bezogen auf die Cashflows, aus denen sich die drei Cashflow-Größen der Kapitalflussrechnung zusammensetzen, keine wesentlichen Bewertungsspielräume vorliegen (vgl. Hitz & Teuteberg, 2013, S. 34). Auch wenn dies ohne Frage richtig ist, ist es wichtig festzuhalten, dass abseits der Bewertungspolitik andere bilanzpolitische Spielräume denkbar sind (vgl. Coenenberg & Meyer, 2003, S. 365). Welche bilanzpolitischen Spielräume abseits der Bewertungspolitik, im Rahmen der Kapitalflussrechnung nach IAS 7, eine Rolle spielen und wie groß deren Bedeutung einzuschätzen ist, soll im Folgenden untersucht werden. Zuerst sollte jedoch der Begriff der Bilanzpolitik näher definiert werden und ein Überblick über die Ziele gegeben werden, die ein Unternehmen mit seiner Bilanzpolitik, insbesondere im Zusammenhang mit der Kapitalflussrechnung, verfolgen kann.

Der Begriff Bilanzpolitik steht im Allgemeinen für die zielorientierte Beeinflussung einzelner Bestandteile des Jahresabschlusses durch das Unternehmen, um eine Verhaltensbeeinflussung des externen Adressaten zu bewirken (vgl. Brösel, 2017, S. 85). Wichtig ist an dieser Stelle zu betonen, dass nach dem allgemeinen Verständnis nur bilanzpolitische Maßnahmen, die im Rahmen der geltenden Gesetze und Rechnungslegungsvorschriften vorgenommen werden, unter diese Definition fallen (vgl. Meyer, 2007, S. 358). Bilanzpolitisches Gestaltungspotential entsteht also nicht etwa dadurch, dass geltende Vorschriften mutwillig missachtet werden, sondern liegt vielmehr darin begründet, dass bestimmte Rechnungslegungsvorschriften explizite oder implizite Wahlrechte gewähren, die vom Unternehmen genutzt werden (vgl. Brösel, 2017, S. 86).

Die Ziele, die ein Unternehmen mit seinen bilanzpolitischen Maßnahmen verfolgt, orientieren sich grundsätzlich an den übergeordneten Unternehmenszielen (vgl. Meyer, 2007, S. 358). Dabei kann eine grobe Unterscheidung vorgenommen werden in Maßnahmen, die darauf abzielen die wirschaftliche Lage des Unternehmens möglichst optimistisch darzustellen und solche Maßnahmen, die eine eher pessimistische Darstellung bezwecken sollen (vgl. Brösel, 2017, S. 90f.). Eine Kombination dieser beiden Ansätze stellt die Erfolgsglättung dar: Verfolgt ein Unternehmen das Ziel der Erfolgsglättung, so führt es in erfolgreichen Perioden bilanzpolitische Maßnahmen durch, welche die wirschaftliche Lage eher pessimistisch darstellen und in schwächeren Perioden solche Maßnahmen, welche die wirschaftliche Lage eher optimistisch darstellen, sodass letztlich größere Schwankungen in der kommunizierten Erfolgslage vermieden werden (vgl. Brösel, 2017, S. 91).

$$\text{Brutto-FCF} = \begin{array}{c}\text{Cashflow aus betrieblicher Tätigkeit}\\ + \text{Cashflow aus Investitionstätigkeit}\end{array}$$

$$\text{Netto-FCF} = \begin{array}{c}\text{Cashflow aus betrieblicher Tätigkeit}\\ + \text{Cashflow aus Investitionstätigkeit}\\ + \text{Veränderungen des Fremdkapitals}\\ - \text{Zinsauszahlungen}\end{array}$$

Abbildung 8: Free-Cashflow-Größen für die Unternehmensbewertung

vgl. Schultze & Meyer, 2005, S. 298

Im Kontext der Kapitalflussrechnung werden bilanzpolitische Maßnahmen in erster Linie durchgeführt, um dem Adressaten ein möglichst optimistisches Bild der wirtschaftlichen Lage zu vermitteln (vgl. Meyer, 2007, S. 367). Als Minimalziel gilt dabei die Signalisierung einer jederzeitigen Zahlungsfähigkeit und daran anschließend die Einhaltung des finanziellen Gleichgewichtes durch das Unternehmen (vgl. Meyer, 2007, S. 377). Außerdem sollen einzelne Cashflow-Größen der Kapitalflussrechnung nach Möglichkeit positiv beeinflusst werden, sodass der Adressat das Unternehmen als möglichst finanzstark wahrnimmt (vgl. Meyer, 2007, S. 377). Für die Umsetzung des letztgenannten Ziels kommt insbesondere der Cashflow aus betrieblicher Tätigkeit sowie der Brutto-FCF, vereinfacht gebildet als Summe der Cashflows aus betrieblicher Tätigkeit und Investitionstätigkeit, in Frage (vgl. Meyer, 2007, S. 377). Als ein wesentliches Ziel der Bilanzpolitik im Kontext der Kapitalflussrechnung sollen diese beiden Cashflow-Größen mittels bilanzpolitischer Maßnahmen so beeinflusst werden, dass sie entweder einen stetigen Anstieg verzeichnen, oder möglichst konstant auf einem erhöhten Niveau verbleiben (vgl. Coenenberg & Meyer, 2003, S. 364). Letzteres zeigt insbesondere, dass auch die Erfolgsglättung ein bilanzpolitisches Ziel im Kontext der Kapitalflussrechnung darstellen kann: Durch die Vermeidung stärkerer Schwankungen dieser wichtigen Cashflow-Größen, möchte das Unternehmen vor allem der Entstehung von Nervosität unter gegenwärtigen und potentiellen Kapitalgebern entgegen wirken (vgl. Coenenberg & Meyer, 2003, S. 364).

Zusammenfassend lassen sich somit der Cashflow aus betrieblicher Tätigkeit sowie der Brutto-FCF als vorrangige Zielgrößen der Bilanzpolitik im Kontext der Kapitalflussrechnung definieren (vgl. Meyer, 2007, S. 380). In den anschließenden beiden Abschnitten sollen nun die konkreten bilanzpolitischen Maßnahmen erläutert werden, die ein Unternehmen im Rahmen der Kapitalflussrechnung ergreifen kann, um seine bilanzpolitischen Ziele zu erreichen.

4.2. Sachverhaltsgestaltende Bilanzpolitik im Rahmen der Kapitalflussrechnung

Bilanzpolitische Maßnahmen lassen sich grundsätzlich (vgl. Lachnit & Müller, 2017, S. 78f.), sowie auch im Kontext der Kapitalflussrechnung (vgl. Meyer, 2007, S. 382), unterscheiden in Sachverhaltsgestaltung und Sachverhaltsabbildung.

Sachverhaltsgestaltende Maßnahmen innerhalb der Kapitalflussrechnung sind stets materieller Art (vgl. Abb. 9), insofern als dass sie dem Unternehmen ermöglichen, die Höhe der in der Kapitalflussrechnung ausgewiesenen Cashflow-Größen zu beeinflussen (vgl. Coenenberg & Meyer, 2003, S. 357). Darüber hinaus steuert ein Unternehmen durch sachverhaltsgestaltende Maßnahmen, anders als durch sachverhaltsabbildende Maßnahmen (vgl. Kapitel 4.3), die Höhe der realen Ein- und Auszahlungen die in einer Periode anfallen und nicht lediglich Form oder Höhe des Ausweises dieser Ein- und Auszahlungen in der Kapitalflussrechnung (vgl. Coenenberg & Meyer, 2003, S. 365f.). Unterscheiden lässt sich die sachverhaltsgestaltende Bilanzpolitik indes in Terminpolitik und Transaktionspolitik (vgl. Eiselt & Müller, 2014, S. 111). Erstere umfasst die Steuerung des Zeitpunktes der zahlungswirksamen Erfassung eines Geschäftsvorfalles, etwa durch Vorziehen des Zahlungszeitpunktes aus einer Folgeperiode, oder durch Aufschieben des Zahlungszeitpunktes in eine Folgeperiode (vgl. Meyer, 2007, S. 382). Als Beispiel für diese Art der Bilanzpolitik wäre das Vorziehen einer Tilgung von Verbindlichkeiten, die eigentlich erst in einer Folgeperiode erfolgen müsste, zu nennen (vgl. Meyer, 2007, S. 382). Da vorgezogene Zahlungsflüsse in der Zukunft nicht erneut auftreten und aufgeschobene Zahlungsflüsse entsprechend in Zukunft nachgeholt werden ist indes zu berücksichtigen, dass Terminpolitik immer eine gegensätzliche Zahlungswirkung in einer folgenden Periode nach sich zieht (vgl. Meyer, 2007, S. 382).

Die Transaktionspolitik umfasst insbesondere die bilanzpolitisch motivierte Durchführung oder Unterlassung zahlungswirksamer Transaktionen, sowie die Modifizierung von Geschäftsvorfällen dahingehend, dass diese zwar wie geplant

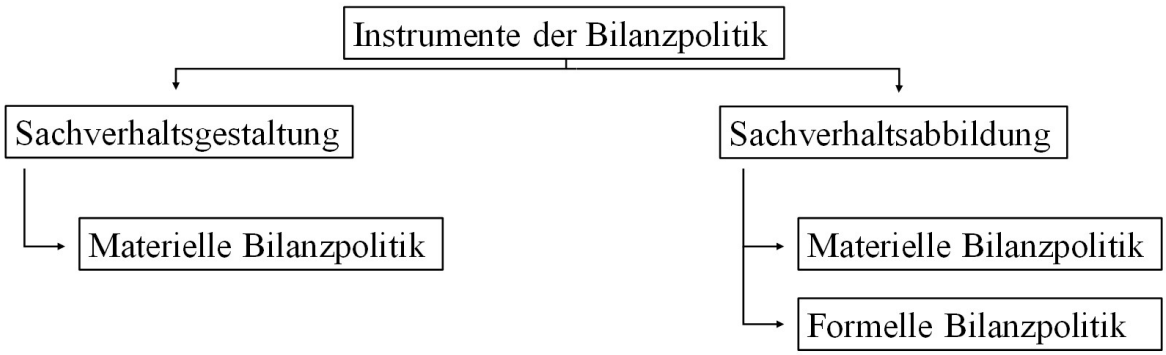

Abbildung 9: Instrumente der Bilanzpolitik

vgl. Meyer, 2007, S. 282

durchgeführt, aber in der betrachteten Periode nicht zahlungswirksam erfasst werden (vgl. Meyer, 2007, S. 382f.). Um Letzteres zu erreichen, können die Merkmale eines Geschäftsvorfalles auf unterschiedliche Weise modifiziert werden: Eine Möglichkeit wäre beispielsweise einen Vermögenswert nicht in bar, sondern durch Schuldübernahme zu erwerben, sodass in der betrachteten Periode keine Auszahlungen in der Kapitalflussrechnung erfasst werden (vgl. Lüdenbach et al., 2020, § 3, Rn. 176).

Durch sachverhaltsgestaltende Maßnahmen lässt sich also die Zahlungswirksamkeit von Geschäftsvorfällen einer Periode gezielt beeinflussen. Zum einen kann ein Unternehmen beeinflussen, ob ein Geschäftsvorfall überhaupt zahlungswirksam erfasst wird und zum anderen, falls ja, in welcher Periode der Geschäftsvorfall zahlungswirksam erfasst wird. Sachverhaltsgestaltende Maßnahmen beeinflussen die Höhe der in der Kapitalflussrechnung ausgewiesenen Cashflow-Größen dabei weitestgehend unbemerkt, da auf Grundlage des Jahresabschlusses für den externen Adressaten keine Möglichkeit besteht beispielsweise nachzuvollziehen, welche Einzahlungen aus originären Geschäftsvorfällen der betrachteten Periode resultieren und welche aus Geschäftsvorfällen einer zukünftigen Periode vorgezogen wurden. Aus diesem Grund ist sachverhaltsgestaltende Bilanzpolitik auf Basis des Jahresabschlusses kaum als solche identifizierbar (vgl. Brösel, 2017, S. 96) und entsprechend liegen auch keine relevanten empirischen Studien vor, die untersuchen, in welchem Maße sie in der Praxis im Rahmen der Kapitalflussrechnung zum Einsatz kommt. Insofern wird der Fokus für die Untersuchung, in welchem Maße bilanzpolitische Spielräume im Kontext der Kapitalflussrechnung nach IAS 7 genutzt werden, im folgenden Abschnitt auf die sachverhaltsabbildenden Maßnahmen gelegt werden.

4.3. Sachverhaltsabbildende Bilanzpolitik im Rahmen der Kapitalflussrechnung

Unter sachverhaltsabbildender Bilanzpolitik werden bilanzpolitische Maßnahmen subsumiert, mit denen sich Form und Höhe des Ausweises von Ein- und Auszahlungen in der Kapitalflussrechnung nach IAS 7, sowie die grundlegende Darstellungsform der Kapitalflussrechnung beeinflussen lassen (vgl. Coenenberg & Meyer, 2003, S. 358). Sachverhaltsabbildende Maßnahmen können sowohl materieller Art als auch formeller Art sein (vgl. Coenenberg & Meyer, 2003, S. 360f.). Im Rahmen materieller Abbildungspolitik wird die Höhe der Cashflow-Größen der Kapitalflussrechnung beeinflusst, insbesondere durch die Zuordnungswahlrechte (vgl. Kapitel 2.2) für Zins-, Dividenden- und Ertragssteuerzahlungen sowie durch den Ermessensspielraum (vgl. Kapitel 2.1), der dem Unternehmen hinsichtlich der Abgrenzung des Finanzmittelfonds gewährt wird (vgl. Meyer, 2007, S. 384). Im Folgenden soll zuerst auf die expliziten Zuordnungswahlrechte eingegangen werden, die der IAS 7 gewährt: Gezahlte und erhaltene Zinsen können demnach, sofern betriebswirtschaftlich begründbar, sowohl dem Cashflow aus betrieblicher Tätigkeit, als auch den Cashflows aus Investitionstätigkeit und Finanzierungstätigkeit zugeordnet werden (vgl. IAS 7.33). Gleiches gilt für erhaltene Dividendenzahlungen (vgl. IAS 7.33). Gezahlte Dividenden können sowohl dem Cashflow aus betrieblicher Tätigkeit, als auch dem Cashflow aus Finanzierungstätigkeit zugerechnet werden (vgl. IAS 7.34). Zudem werden Ertragssteuerzahlungen grundsätzlich dem Cashflow aus betrieblicher Tätigkeit zugeordnet (vgl. IAS 7.35). Zwar sollen gemäß IAS 7.35 Ertragssteuerzahlungen, sofern die Zuordnung zu einer bestimmten Investitions- oder Finanzierungstätigkeit möglich ist, auch den entsprechenden Cashflow-Bereichen der Kapitalflussrechnung zugerechnet werden, dies ist allerdings häufig schwer umsetzbar (vgl. IAS 7.36). Darüber hinaus müssen insbesondere die genannten Wahlrechte für Zins- und Dividendenzahlungen nicht für Ein- und Auszahlungen einheitlich ausgeübt werden: Es besteht beispielsweise die Möglichkeit Zinseinzahlungen dem Cashflow aus betrieblicher Tätigkeit zuzuordnen, während die Zinsauszahlungen dem Cashflow aus Finanzierungstätigkeit zugeordnet werden (vgl. Hüttche & Int-Veen, 2008, S. 349).

Diese weitreichenden Zuordnungswahlrechte hinsichtlich Zins-, Dividenden- und Ertragssteuerzahlungen deuten

auf ein signifikantes bilanzpolitisches Potential hin. Wie in Kapitel 4.1 erläutert, stellt der Ausweis eines stetigen, (moderat) steigenden Brutto-FCF beziehungsweise Cashflows aus betrieblicher Tätigkeit ein wesentliches Ziel der Bilanzpolitik im Kontext der Kapitalflussrechnung dar. Vor diesem Hintergrund wäre zu erwarten, dass ein Unternehmen, sofern betriebswirtschaftlich begründbar, seine Einzahlungen aus Zins und Dividende im Cashflow aus betrieblicher Tätigkeit ausweist und seine Auszahlungen aus Zins, Dividende und falls umsetzbar auch aus Ertragssteuer, dem Cashflow aus Finanzierungstätigkeit zuordnet, um sowohl den Cashflow aus betrieblicher Tätigkeit als auch den Brutto-FCF möglichst hoch darzustellen.

Empirische Studien, die gerade die Nutzung dieser Zuordnungswahlrechte in der unternehmerischen Praxis untersuchen, liegen in größerer Zahl vor: In einer empirischen Untersuchung der IFRS-Konzernabschlüsse von Unternehmen aus den großen deutschen Aktienindizes zeigen Haller, Groß und Rauscher (2014, S. 301), dass alleine bei der Zuordnung von Zins- und Dividendenzahlungen im Durchschnitt über die Zuweisung eines Zahlungsmittelbetrages entschieden wird, der dem Vierfachen der durchschnittlichen zahlungswirksamen Änderung des Finanzmittelfonds entspricht. Dies unterstreicht noch einmal das bilanzpolitische Potential, das aus den Zuordnungswahlrechten resultiert. Weiterhin geht aus der Studie hervor, dass die gewährten Wahlrechte in der Praxis auch tatsächlich divergent genutzt werden, was sich unter anderem darin zeigt, dass Zins- und Dividendeneinzahlungen in der Stichprobe allen drei Cashflow-Bereichen zugeordnet wurden (vgl. Haller et al., 2014, S. 302). Stark unterschiedliche Praktiken hinsichtlich des Ausweises von Zinseinzahlungen beobachten auch Kaiser und Bierwulf (2011, S. 10), sowie Bösser, Pilhofer und Lessel (2013, S. 363f.). Die vermutete Ausnutzung der Zuordnungswahlrechte für den Ausweis eines höheren Brutto-FCF beziehungswiese Cashflows aus betrieblicher Tätigkeit durch unterschiedliche Zuordnung von Ein- und Auszahlungen, lässt sich indes lediglich mit Blick auf Dividendenzahlungen feststellen: Während Zinseinzahlungen und -auszahlungen selten getrennt voneinander in unterschiedlichen Cashflow-Bereichen ausgewiesen werden, werden Dividendeneinzahlungen tendenziell im Cashflow aus betrieblicher Tätigkeit ausgewiesen, während Dividendenauszahlungen mehrheitlich dem Cashflow aus Finanzierungstätigkeit zugerechnet werden (vgl. Haller et al., 2014, S. 302). Zuletzt kann noch festgestellt werden, dass der Ausweis von Ertragssteuerzahlungen im Cashflow aus Investitionstätigkeit oder Finanzierungstätigkeit keinerlei praktische Relevanz besitzt: Sämtliche Unternehmen in der Studie von Hüttche und Int-Veen (2008, S. 350) weisen ihre Ertragssteuerzahlungen im Cashflow aus betrieblicher Tätigkeit aus, was mit der erwähnten Zuordnungsschwierigkeit dieser Zahlungen zu einer konkreten Investitions- oder Finanzierungstätigkeit zusammenhängen dürfte.

Neben den erläuterten Zuordnungswahlrechten ist auch die Abgrenzung des Finanzmittelfonds, als weiterer Bestandteil der materiellen Abbildungspolitik, ein beeinflussender Faktor für die Höhe der ausgewiesenen Cashflow-Größen in der Kapitalflussrechnung. Da die Signalisierung jederzeitiger Zahlungsfähigkeit als ein Minimalziel der Bilanzpolitik im Kontext der Kapitalflussrechnung gilt (vgl. Kapitel 4.1), wird ein Unternehmen bilanzpolitische Maßnahmen so durchführen, dass der Finanzmittelfonds eine Mindesthöhe nicht unterschreitet und im Zweifel höher ausfällt (vgl. Meyer, 2007, S. 450). Wie in Kapitel 2.1 erwähnt, bestehen sowohl beim Einbezug von Kontokorrentkrediten als auch bei der Abgrenzung der Zahlungsmitteläquivalente als Bestandteil des Finanzmittelfonds Ermessensspielräume: Beispielsweise dürfen Zahlungsmitteläquivalente gemäß IAS 7.7 ausschließlich mit dem Ziel gehalten werden, kurzfristige Zahlungsverpflichtungen zu erfüllen. Das Ziel mit dem ein Zahlungsmitteläquivalent gehaltenen wird kann allerdings durch den externen Adressaten nicht nachvollzogen werden (vgl. Hüttche & Int-Veen, 2008, S. 351f.), sodass sich faktisch ein Ermessensspielraum ergibt.

In empirischen Studien zeigt sich zum einen, dass Unternehmen, konsistent mit dem bilanzpolitischen Ziel einen tendenziell höheren Finanzmittelfonds auszuweisen, Kontokorrentkredite nur selten in den Finanzmittelfonds mit einbeziehen (vgl. Hüttche & Int-Veen, 2008, S. 352f.). Der diesbezüglich vorliegende Ermessensspielraum wird demnach nicht häufig genutzt. Zum anderen zeigt sich, dass eine Vielzahl unterschiedlicher Posten im Finanzmittelfonds der betrachteten Unternehmen berücksichtigt wird, was darauf hindeutet, dass die Ermessensspielräume hinsichtlich der Abgrenzung der Zahlungsmitteläquivalente in der Praxis für einen weitreichenden Einbezug verschiedener Zahlungsmitteläquivalente in den Finanzmittefonds genutzt werden (vgl. Haller et al., 2014, S. 300). Darüber hinaus sind die Informationen, die Unternehmen hinsichtlich der Zusammensetzung des Finanzmittelfonds im Anhang ihres Jahresabschlusses bereit stellen häufig derart unspezifisch, dass der Adressat kaum nachvollziehen kann, welche der angegebenen Posten den Zahlungsmitteln und welche Posten den Zahlungsmitteläquivalenten zuzurechnen sind (vgl. Haller et al., 2014, S. 300). Die Folge ist, dass der Anteil, den die Zahlungsmitteläquivalente am gesamten Finanzmittelfonds ausmachen, häufig nicht genau beziffert werden kann (vgl. Haller et al., 2014, S. 300). Vor diesem Hintergrund besitzen die Ermessensspielräume bezüglich der Abgrenzung der Zahlungsmitteläquivalente eine signifikante bilanzpolitische Bedeutung.

Als zweiter Bestandteil der Abbildungspolitik, zielt die formelle Abbildungspolitik insbesondere auf eine Beeinflussung der Darstellungsform der einzelnen Cashflow-Größen innerhalb der Kapitalflussrechnung ab (vgl. Meyer, 2007, S. 386). Für den Zweck dieser Arbeit, soll der Fokus auf ein bedeutendes formelles Abbildungswahlrecht gelegt werden: Das Wahlrecht zwischen direkter und indirekter Darstellung des Cashflows aus betrieblicher Tätigkeit. Wie bereits in Kapitel 2.2 erwähnt, wird der direkten Darstellung des Cashflows aus betrieblicher Tätigkeit in der Literatur eine höhere Aussagekraft zugesprochen, was vor allem damit zusammenhängt, dass aus der direkten Darstellung die einzelnen operativen Ein- und Auszahlungen direkt entnommen werden können (vgl. Meyer, 2007, S. 388). Diese Ein- und Aus-

zahlungen können von Adressatenseite für Unternehmens- und Zeitvergleiche genutzt werden (vgl. Broome, 2004, S. 18). Die indirekte Darstellung stellt hingegen eine Art Überleitungsrechnung zwischen Periodenergebnis und Cashflow dar (vgl. Kapitel 2.2) und weist operative Ein- und Auszahlungen nicht explizit aus, sodass lediglich aus der direkten Darstellung das Zustandekommen des Cashflows aus betrieblicher Tätigkeit, aus den operativen Ein- und Auszahlungen (im Folgenden auch "Cashflow-Komponenten" genannt), ersichtlich wird (vgl. Krishnan & Largay III, 2000, S. 215). Darauf aufbauend wird häufig als weiterer Vorteil der direkten Darstellung genannt, dass auf Basis der ausgewiesenen Cashflow-Komponenten, eine präzisere Prognose zukünftiger Cashflows aus betrieblicher Tätigkeit möglich ist (vgl. Meyer, 2007, S. 64f.). Nicht vernachlässigt werden soll an dieser Stelle, dass auch Argumente für die Nutzung der indirekten Darstellung existieren: Demnach verdeutlicht die Überleitungsrechnung den Zusammenhang zwischen Periodenergebnis und Cashflow und hebt auf diesem Wege den Einfluss bestimmter Periodenabgrenzungen, insbesondere der Abschreibungen, auf das Periodenergebnis hervor (vgl. Meyer, 2007, S. 65).

In der Praxis ist unter den externen Adressaten des Jahresabschlusses allerdings ebenfalls eine Präferenz für die direkte Darstellung des Cashflows aus betrieblicher Tätigkeit zu erkennen: In einer Befragung unter Finanz- und Kreditanalysten in Australien präferierten über 70% der Befragten die direkte Darstellung und weniger als 5% die indirekte Darstellung (vgl. Jones & Widjaja, 1998, S. 213f.). Darüber hinaus betont das Chartered Financial Analyst (CFA) Institute, ein Zusammenschluss von Finanzanalysten mit Sitz in den Vereinigten Staaten von Amerika (USA), im Rahmen einer Stellungnahme, dass selbst geübte Analysten nicht in der Lage seien, die Cashflow-Komponenten der direkten Darstellung aus der indirekten Darstellung fehlerfrei herzuleiten (vgl. CFA Institute Centre for Financial Market Integrity, 2007, S. 22).

Trotz des offenkundigen Zuspruchs für die direkte Darstellung des Cashflows aus betrieblicher Tätigkeit aus Theorie und Praxis, hat das explizite Darstellungswahlrecht gemäß IAS 7.18 Bestand. Außerdem zeigt sich in empirischen Studien, dass in der Praxis nahezu alle untersuchten Unternehmen das Wahlrecht zur indirekten Darstellung des Cashflows aus betrieblicher Tätigkeit auch tatsächlich ausüben (vgl. Haller et al., 2014, S. 302; Hitz & Teuteberg, 2013, S. 38). In Kapitel 5.4 wird anhand empirischer Studien untersucht werden, ob die Cashflow-Komponenten, die der direkten Darstellung unmittelbar zu entnehmen sind, eine präzisere Prognose zukünftiger Cashflows aus betrieblicher Tätigkeit ermöglichen, als die aggregierte, indirekte Darstellung. Außerdem wird untersucht werden, ob die Herleitung dieser Komponenten aus der indirekten Darstellung tatsächlich nur mit signifikantem Fehler erfolgen kann. Falls sich dies bestätigt wäre gezeigt, dass die direkte Darstellung des Cashflows aus betrieblicher Tätigkeit dem Adressaten des Jahresabschlusses Informationen liefert, die aus der indirekten Darstellung nicht hervor gehen. Dies würde bedeuten, dass ein bilanzierendes Unternehmen, durch die Ausübung des Wahlrechtes zur indirekten Darstellung des Cashflows aus betrieblicher Tätigkeit, die Aussagekraft seiner Kapitalflussrechnung signifikant reduziert.

4.4. Grenzen der Bilanzpolitik

In diesem Abschnitt soll auf wesentliche Faktoren eingegangen werden, welche die Bedeutung bilanzpolitischer Maßnahmen in der Praxis begrenzen. Ein wichtiger Faktor ist die Erkennbarkeit der bilanzpolitischen Maßnahme: Kann der Adressat bei Betrachtung des Jahresabschlusses die angewandte bilanzpolitische Maßnahme identifizieren, so verliert sie an verhaltensbeeinflussender Wirkung (vgl. Meyer, 2007, S. 459). Diese Feststellung ist von besonderer Bedeutung für die Zuordnungswahlrechte hinsichtlich Zins-, Dividenden- und Ertragssteuerzahlungen: Zwar sind auf den ersten Blick, insbesondere durch den asymmetrischen Ausweis von Dividendeneinzahlungen und Dividendenauszahlungen signifikante bilanzpolitische Verzerrungen möglich (vgl. Kapitel 4.3). Wie aber bereits in Kapitel 3.3 ausgeführt, müssen Zins-, Dividenden- und Ertragssteuerzahlungen entweder innerhalb der Ursachenrechnung, oder zumindest im Anhang des Jahresabschlusses separat angegeben werden. Ein geübter Adressat kann, im Rahmen seiner Analyse der Kapitalflussrechnung, die betreffenden Ein- und Auszahlungen also relativ leicht identifizieren und für den Zweck seiner Analyse umordnen (vgl. Eiselt & Müller, 2014, S. 39). Gegenteiliges gilt jedoch hinsichtlich des Ermessensspielraums bei der Abgrenzung des Finanzmittelfonds: Die Informationen die Unternehmen hinsichtlich der Zusammensetzung des Finanzmittelfonds im Anhang ihres Jahresabschlusses bereitstellen, sind in der Regel zu undifferenziert (vgl. Kapitel 4.3), als dass der Adressat die unterschiedlich genutzten Spielräume bezüglich der Abgrenzung des Finanzmittelfonds erkennen, oder gar korrigieren könnte (vgl. Haller et al., 2014, S. 301). Die Tatsache, dass die Erkennbarkeit der bilanzpolitischen Maßnahme zu einem Bedeutungsverlust ebendieser führt, rückt zudem die sachverhaltsgestaltende Bilanzpolitik in den Fokus: Wie ein Kapitel 4.2 festgestellt, ist sachverhaltsgestaltende Bilanzpolitik auf Basis des Jahresabschlusses kaum als solche identifizierbar, sodass ihr auf den ersten Blick eine besondere bilanzpolitische Bedeutung zugeordnet werden könnte. In der Praxis wird die Bedeutung von sachverhaltsgestaltenden Maßnahmen allerdings dadurch begrenzt, dass die Durchführung dieser Maßnahmen häufig mit einem erheblichen Planungs- und Abstimmungsaufwand verbunden ist, sodass der Umsetzungsaufwand schnell den bilanzpolitischen Nutzen der Maßnahme übersteigt (vgl. Meyer, 2007, S. 455).

Darüber hinaus wird die Bedeutung bilanzpolitischer Maßnahmen, zumindest in gewissem Maße, auch durch das Prinzip der Stetigkeit begrenzt. Gemäß dem Prinzip der Stetigkeit müssen auch die bilanzpolitischen Spielräume bezüglich der Kapitalflussrechnung im Zeitablauf konsistent ausgelegt werden (vgl. Lüdenbach et al., 2020, § 3, Rn. 5). Dies erleichtert zwar den Vergleich im Zeitablauf, da beispielsweise Zuordnungen zu den Zahlungsmitteläquivalenten im

Zeitablauf in der Regel beibehalten werden müssen, nicht jedoch den Unternehmensvergleich.

Insgesamt ist also festzuhalten, dass sich die Bedeutung bestimmter bilanzpolitischer Verzerrungen, etwa der Zuordnungswahlrechte, durch einen erfahrenen Adressaten mittels Aufbereitungsmaßnahmen begrenzen lässt. Solche Aufbereitungsmaßnahmen gelangen allerdings schon im Kontext der Fondsabgrenzung an ihre Grenzen. Die Untersuchung zeigt somit, dass bilanzpolitische Spielräume im Kontext der Kapitalflussrechnung nach IAS 7 vorliegen, die insbesondere auch eine Beeinflussung der Höhe der ausgewiesenen Cashflow-Größen ermöglichen. Da der Adressat diesen bilanzpolitischen Verzerrungen nur eingeschränkt durch Aufbereitungsmaßnahmen entgegenwirken kann, ist die Vergleichbarkeit der Kapitalflussrechnung, besonders im Unternehmensvergleich, durchaus signifikant eingeschränkt. Insofern muss der Annahme, die Kapitalflussrechnung sei weitestgehend immun gegenüber bilanzpolitischen Maßnahmen, widersprochen werden, auch wenn sich, anders als beim Periodenergebnis, kaum Bewertungsspielräume bieten. Im folgenden Kapitel soll die Aussagekraft von Cashflow-Größen der Kapitalflussrechnung weitergehend untersucht werden indem die Annahme, das Periodenergebnis sei durch Berücksichtigung betriebswirtschaftlich notwendiger Periodenabgrenzungen konzeptionell besser für die Abbildung des Unternehmenserfolges geeignet als etwaige Cashflow-Größen, untersucht wird.

5. Vergleich der Aussagekraft von Cashflow und Periodenergebnis als Indikatoren des Unternehmenserfolges

5.1. Periodenabgrenzungen als konzeptioneller Unterschied zwischen Periodenergebnis und Cashflow

Um die Aussagekraft von Periodenergebnis und Cashflow, mit Blick auf die Kapitalflussrechnung nach IAS 7, vergleichend betrachten zu können, muss vorab beleuchtet werden, wie sich das Periodenergebnis konzeptionell von den Cashflow-Größen aus der Kapitalflussrechnung unterscheidet: Als zentraler Unterschied ist die Berücksichtigung von Periodenabgrenzungen zu nennen (vgl. Dechow, Kothari & Watts, 1998, S. 152). Während die Cashflow-Größen der Ursachenrechnung nach IAS 7, wie in Kapitel 2 dargestellt, ausschließlich Cashflows umfassen, setzt sich das Periodenergebnis der Gesamtergebnisrechnung sowohl aus Cashflows als auch aus zahlungsunwirksamen Periodenabgrenzungen zusammen (vgl. Farshadfar & Monem, 2013, S. 1061). Die Periodenabgrenzungen lassen sich dabei in die folgenden, wesentlichen Komponenten kategorisieren: Veränderungen von Forderungen und Verbindlichkeiten, Veränderungen des Vorratsvermögens, Abschreibungen auf materielle und immaterielle Vermögenswerte sowie Bildung und Auflösung von Rückstellungen (vgl. Barth, Cram & Nelson, 2001, S. 28). Aufgabe dieser Periodenabgrenzungen ist es, die volatil auftretenden Cashflows abzugrenzen und den jeweiligen Perioden so zuzuordnen, dass der Unternehmenserfolg einer Periode korrekt abgebildet wird (vgl. Kühnberger & Thurmann, 2015, S. 35). Da es also gerade Ziel der Rechnungslegung mittels Periodenabgrenzungen ist, eine periodische Erfolgsgröße zu ermitteln, die in ihrer Aussagekraft der reinen Betrachtung von Cashflows überlegen ist, liegt die Behauptung, das Periodenergebnis sei grundsätzlich besser für die Abbildung des Unternehmenserfolges geeignet, nahe (vgl. Gebhardt & Mansch, 2012, S. 130). Diese Ansicht wurde bereits im Jahr 1978 durch das in den USA tätige Financial Accounting Standards Board (FASB) vertreten (vgl. FASB, 1978, S. 13). In der Literatur finden sich darüber hinaus zahlreiche Beiträge, die diese Einschätzung argumentierend unterstützen: Demnach bildet das Periodenergebnis den Unternehmenserfolg besser ab, da nicht wertrelevante Cashflows der betrachteten Periode abgrenzt und nicht berücksichtigt werden, während wertrelevante Cashflows anderer Perioden miteinbezogen werden (vgl. Liu, Nissim & Thomas, 2007, S. 58). Die Schwäche von Cashflow-Größen liegt demnach in erster Linie darin begründet, dass zeitliche und inhaltliche Zuordnungsprobleme hinsichtlich der Ein- und Auszahlungen bestehen, da betriebswirtschaftlich notwendige Periodenabgrenzungen nicht vorgenommen werden (vgl. Dechow, 1994, S. 4).

Allerdings ist zu berücksichtigen, dass insbesondere in der Praxis, trotz der oben erläuterten Argumentationslinie, den Cashflow-Größen eine hohe Bedeutung für die Bemessung des Unternehmenserfolges zugerechnet wird: Häufig wird betont, dass die im Periodenergebnis enthaltenen Periodenabgrenzungen eher bilanzpolitischen Maßnahmen unterliegen können, da sie nicht wie die Zahlungsflüsse hinsichtlich Höhe und Zeitpunkt des Anfalls grundsätzlich determiniert sind (vgl. Gebhardt & Mansch, 2012, S. 131). Einschränkend muss an dieser Stelle auf die Analyse der Bilanzpolitik im Rahmen der Kapitalflussrechnung aus dem vorherigen Kapitel verwiesen werden, die gezeigt hat, dass Cashflow-Größen sehr wohl auch signifikanten bilanzpolitischen Maßnahmen unterliegen können. Gleichwohl ist es richtig, dass gerade Bewertungsspielräume im Rahmen der Periodenabgrenzungen das Periodenergebnis maßgeblich verzerren können, die im Kontext der Kapitalflussrechnung keine Rolle spielen.

Hinsichtlich der Nützlichkeit von Periodenabgrenzungen, für die Ermittlung eines möglichst aussagekräftigen Periodenergebnisses liegen somit unterschiedliche Ansichten vor. Um also zu untersuchen, ob das Periodenergebnis tatsächlich konzeptionell besser für die Abbildung des Unternehmenserfolges geeignet ist als etwaige Cashflow-Größen, muss letztlich die folgende Frage beantwortet werden: Erhöhen Periodenabgrenzungen als wesentliche Differenz zwischen Cashflow-Größe und dem Periodenergebnis tatsächlich die Aussagekraft des Periodenergebnisses, oder verzerren bewertungspolitische Maßnahmen im Kontext der Periodenabgrenzungen das Periodenergebnis so stark, dass stattdessen Cashflow-Größen, die zumindest keinen wesentlichen Bewertungsspielräumen unterliegen, den Unternehmenserfolg besser abbilden?

Da das FASB in seinem Statement 1978 die Behaup-

tung, das Periodenergebnis bilde den Unternehmenserfolg besser ab als etwaige Cashflow-Größen, ohne eine empirische Grundlage aufstellte, wurden in den folgenden Jahren diverse empirische Untersuchungen durchgeführt, um die aufgestellte These zu überprüfen. Im kommenden Abschnitt soll diese Literatur genauer untersucht werden.

5.2. Vergleichende Betrachtung der Wertrelevanz von Cashflow und Periodenergebnis anhand von Aktienrenditen

Im Wesentlichen wird die Aussagekraft von Cashflow-Größe und Periodenergebnis in empirischen Studien auf Basis zweier unterschiedlicher Ansätze untersucht: Zum einen wird geprüft, ob die gegenwärtige Cashflow-Größe oder das gegenwärtige Periodenergebnis eine genauere Prognose zukünftiger Cashflow-Größen ermöglicht. Zum anderen wird die Wertrelevanz von Cashflow-Größe und Periodenergebnis vergleichend betrachtet, indem untersucht wird welche Größe in stärkerem Zusammenhang mit der Aktienrendite eines Unternehmens steht. In diesem Abschnitt wird zuerst die Literatur analysiert die letzteren Ansatz verfolgt, bevor im folgenden Abschnitt die Prognosekraft von Cashflow-Größe und Periodenergebnis betrachtet wird. Generell wird im Rahmen aller Untersuchungen ein operativer Cashflow als Cashflow-Größe definiert, der allerdings nicht zwangsläufig dem Cashflow aus betrieblicher Tätigkeit nach IAS 7 entsprechen muss: Da ein Großteil der untersuchten, empirischen Studien in den USA durchgeführt wurde, beziehen sich diese Studien entsprechend überwiegend auf die operative Cashflow-Größe aus der Kapitalflussrechnung nach den United States Generally Accepted Accounting Principles (US-GAAP). Auch wenn die Übertragbarkeit der folgenden Erkenntnisse auf eine Kapitalflussrechnung nach IAS 7 deshalb kritisch hinterfragt werden muss, ist die Untersuchung der Literatur dennoch sinnvoll, da sich die Kapitalflussrechnung nach IAS 7 konzeptionell erkennbar an der Kapitalflussrechnung nach US-GAAP orientiert (vgl. Meyer, 2007, S. 112). Darüber hinaus prüfen beispielsweise Nallareddy, Sethuraman und Venkatachalam (2020, S. 18f.) in ihrer Studie explizit die Übertragbarkeit ihrer Ergebnisse auf eine Stichprobe die Unternehmen enthält, die ihre Kapitalflussrechnung nach IAS 7 aufstellen und finden keine signifikanten Abweichungen. Insofern kann für den Zweck dieser Arbeit davon ausgegangen werden, dass die folgenden Erkenntnisse auch für die Aussagekraft der Kapitalflussrechnung nach IAS 7 von Bedeutung sind.

Die Wertrelevanz von Periodenergebnis und operativem Cashflow wird im Folgenden verglichen, indem untersucht wird ob Cashflow oder Periodenergebnis einen stärkeren Zusammenhang zur Aktienrendite aufweisen. Die Annahme die diesem Vorgehen zu Grunde liegt ist, dass auf einem effizienten Aktienmarkt in der Aktienrendite alle wertrelevanten Informationen bezüglich des Unternehmens eingepreist sind und somit die Aktienrendite als Benchmark für den Unternehmenserfolg genutzt werden kann (vgl. Dechow, 1994, S. 12). Als Aktienrendite wird dabei die Summe aus Kursveränderung und Dividendenauszahlung auf einem bestimmten Zeitintervall definiert (vgl. beispielhaft Charitou, Clubb & Andreou, 2000, S. 6). Insgesamt zeigt sich in den betrachteten empirischen Studien, dass über alle untersuchten Intervalllängen hinweg, das Periodenergebnis einen relativ größeren Zusammenhang mit der Aktienrendite aufweist, als die operativen Cashflow-Größen (vgl. Dechow, 1994, S. 21). Auch eine jüngere Studie, basierend auf Daten aus dem Vereinigten Königreich (UK), bestätigt die relativ höhere Wertrelevanz des Periodenergebnisses (vgl. Mostafa, 2014, S. 654). Dabei fällt die Wertrelevanz des operativen Cashflows geringer aus, desto länger der Geschäftszyklus des Unternehmens ist und desto größer entsprechend die Veränderungen in Forderungen, kurzfristigen Verbindlichkeiten und Vorräten sind (vgl. Dechow, 1994, S. 31). Mit zunehmender Relevanz der Periodenabgrenzungen sinkt also die Wertrelevanz des Cashflows, während die Wertrelevanz des Periodenergebnisses unverändert bleibt, was darauf hindeutet, dass die Periodenabgrenzungen tatsächlich die Aussagekraft des Periodenergebnisses gegenüber der Aussagekraft des Cashflows erhöhen (vgl. Dechow, 1994, S. 31).

Gleichwohl zeigen Cheng et al. in ihrer Studie aus dem Jahr 1996, dass insbesondere dann wenn das Periodenergebnis Sondereffekte beinhaltet, die Wertrelevanz des Periodenergebnisses eingeschränkt sein kann (vgl. Cheng, Liu & Schaefer, 1996, S. 180). Zu solchen Sondereffekten zählen beispielsweise Verluste aus Restrukturierung oder auch Einmaleffekte bei Änderung der Rechnungslegungsstandards (vgl. Cheng et al., 1996, S. 177). In solchen Szenarien ist der operative Cashflow von erhöhter Wertrelevanz (vgl. Cheng et al., 1996, S. 180). Zusammenfassend ist festzuhalten, dass sich auf Basis empirischer Wertrelevanz-Studien die Annahme, dass der Unternehmenserfolg besser durch das Periodenergebnis als durch Cashflow-Größen abgebildet wird, bestätigt. Darüber hinaus scheinen gerade die Periodenabgrenzungen, als konzeptioneller Unterschied zwischen Cashflow und Periodenergebnis, die Aussagekraft des Periodenergebnisses zu erhöhen. Der Cashflow kann demnach eher als sekundäre, ergänzende Erfolgsgröße genutzt werden, die betrachtet werden sollte, wenn die Aussagekraft des Periodenergebnisses durch Sondereffekte beschränkt wird. Im folgenden Abschnitt wird überprüft, ob sich die relativ höhere Aussagekraft des Periodenergebnisses auch beim Vergleich der Prognosekraft von Cashflow und Periodenergebnis zeigt.

5.3. Vergleichende Betrachtung der Prognosekraft von Cashflow und Periodenergebnis

Wie in Kapitel 3.1 erläutert, ist ein wichtiges Ziel der Kapitalflussrechnung und der darin enthaltenen Cashflow-Größen, die Prognose zukünftiger Cashflows zu ermöglichen. Diese prognostizierten Cashflows bilden als wichtige Erfolgsgröße eines Unternehmens beispielsweise die Grundlage für die Unternehmensbewertung mittels DCF-Verfahren (vgl. Kapitel 3.3). Vor diesem Hintergrund existiert eine Vielzahl empirischer Studien die untersuchen, ob der gegenwärtige operative Cashflow oder das gegenwärtige Periodenergebnis als Inputfaktor eine präzisere Prognose zukünftiger operativer Cashflows ermöglicht. Die betrachteten Studien sind dabei

immer ähnlich aufgebaut: Mit Hilfe einer Regressionsanalyse wird, basierend auf operativem Cashflow und Periodenergebnis als Inputfaktoren, der zukünftige operative Cashflow prognostiziert (vgl. beispielhaft Greenberg, Johnson & Ramesh, 1986, S. 270). Hinsichtlich ihrer Resultate divergieren die Studien allerdings signifikant, sodass bis heute die zu Grunde liegende Frage, welche der beiden Rechnungslegungsgrößen eine präzisere Cashflow-Prognose ermöglicht, nicht abschließend geklärt ist: Zum einen kommt eine Reihe von Studien zu dem Ergebnis, das Periodenergebnis sei besser als Inputfaktor für die Cashflow-Prognose geeignet (vgl. Dechow et al., 1998, S. 152; Greenberg et al., 1986, S. 271). Zum anderen existieren Studien, die zu genau dem gegenteiligen Ergebnis gelangen und dem gegenwärtigen Cashflow eine relativ höhere Prognosekraft zusprechen (vgl. Finger, 1994, S. 219; Nallareddy et al., 2020, S. 6). Aus diesem Grund ist es notwendig, die verschiedenen Studien auf konzeptionelle Unterschiede zu untersuchen, die diese abweichenden Ergebnisse erklären können. Gemäß Nallareddy et al. stellt die unterschiedliche Bemessung der operativen Cashflow-Größe als Inputfaktor der Regressionsanalyse, den zentralen Unterschied zwischen den verschiedenen Studien dar und ist somit eine Hauptursache für die unterschiedlichen Ergebnisse (vgl. Nallareddy et al., 2020, S. 6). Konsistent mit dieser Hypothese leiten Greenberg et al. (1986, S. 269) genauso wie Dechow et al. (1998, S. 146) ihre operative Cashflow-Größe indirekt, auf Basis von Bilanz und Gesamtergebnisrechnung her, während Nallareddy et al. (2020, S. 4) ihre operative Cashflow-Größe direkt der Kapitalflussrechnung entnehmen. Um ihre Hypothese weiter zu stützen, führen Nallareddy et al. ihre empirische Untersuchung ein weiteres Mal auf Basis einer aus Bilanz und Gesamtergebnisrechnung hergeleiteten operativen Cashflow-Größe durch und kommen, anders als zuvor, tatsächlich zu dem gegensätzlichen Ergebnis, das Periodenergebnis habe eine relativ größere Prognosekraft als der operative Cashflow (vgl. Nallareddy et al., 2020, S. 6). Die relativ geringere Prognosekraft des operativen Cashflows in vorherigen Studien könnte also vor allem daraus resultieren, dass die Herleitung eines operativen Cashflows aus Bilanz und Gesamtergebnisrechnung fehleranfällig ist (vgl. Hribar & Collins, 2002, S. 115).

Darüber hinaus zeigen Nallareddy et al. (2020, S. 15f.), dass die Prognosekraft des operativen Cashflows zunimmt, desto kürzer der Geschäftszyklus des Unternehmens ist und desto geringer entsprechend die Veränderungen in Forderungen, kurzfristigen Verbindlichkeiten und Vorräten sind. Dieses Ergebnis ist konsistent mit der im vorherigen Abschnitt erlangten Erkenntnis, dass mit zunehmender (sinkender) Relevanz der Periodenabgrenzungen, die Aussagekraft des Cashflows sinkt (steigt).

Insgesamt lässt sich auf Basis der vorliegenden Studien nicht abschließend beurteilen, ob dem operativen Cashflow oder dem Periodenergebnis eine höhere Aussagekraft bei der Prognose zukünftiger Cashflows zukommt. Die unterschiedlichen Studienergebnisse lassen sich aber vor allem durch die abweichende Bemessung der operativen Cashflow-Größe erklären.

5.4. Erhöhung der Prognosekraft durch Disaggregation des operativen Cashflows

Aufbauend auf der im vorherigen Abschnitt untersuchten Literatur, die sich mit der Prognosekraft von Cashflow-Größen auseinander setzt, existieren Studien die prüfen, ob sich durch Disaggregation, das heißt durch Aufschlüsselung der operativen Cashflow-Größe in ihre Komponenten, eine präzisere Prognose zukünftiger operativer Cashflows ermitteln lässt. Diese Forschungsfrage ist insbesondere deshalb von Relevanz, da der Adressat die Cashflow-Komponenten, bestehend aus den verschiedenen operativen Ein- und Auszahlungen, der direkten Darstellung des Cashflows aus betrieblicher Tätigkeit zumeist entnehmen kann, während sie aus der indirekten Darstellung nicht hervorgehen (vgl. Kapitel 2.2). Sofern anhand empirischer Studien gezeigt würde, dass auf Basis der Cashflow-Komponenten eine präzisere Prognose gebildet werden kann und eine Herleitung dieser Komponenten aus der indirekten Darstellung nicht fehlerfrei möglich ist, wäre nachgewiesen, dass die direkte Darstellung des Cashflows aus betrieblicher Tätigkeit hinsichtlich ihrer Prognosekraft und somit ihrer Aussagekraft, der indirekten Darstellung überlegen ist. Dieses Resultat legt die Diskussion in Kapitel 4.3 bereits nahe.

Wichtig ist zuerst festzuhalten, in welche wesentlichen Komponenten die operative Cashflow-Größe in den betrachteten Studien unterteilt wird. Sowohl Cheng und Hollie (2008, S. 34) als auch Khansalar und Namazi (2017, S. 467) definieren Umsatzeinzahlungen sowie zahlungswirksame Umsatz- und Betriebskosten als Kern-Komponenten des operativen Cashflows. Zinszahlungen werden genauso wie Steuerzahlungen den sonstigen Komponenten zugeordnet. Trotz der vorhandenen konzeptionellen Ähnlichkeit lohnt sich eine zusätzliche Betrachtung der Studie von Khansalar und Namazi, da diese neben Unternehmen aus den USA auch Unternehmen aus dem UK in die Untersuchung mit einbezieht (vgl. Khansalar & Namazi, 2017, S. 472). Die wesentlichen Ergebnisse der Studien sind dabei identisch: Grundsätzlich sind die Kern-Komponenten des operativen Cashflows zeitbeständiger als die sonstigen Komponenten, während Steuerzahlungen die geringste Beständigkeit aufweisen (vgl. Cheng & Hollie, 2008, S. 43; Khansalar & Namazi, 2017, S. 472). Somit sind die Kern-Komponenten grundsätzlich besser für die Prognose geeignet und können gezielt eingesetzt werden, um die Prognosekraft des operativen Cashflows zu erhöhen (vgl. Khansalar & Namazi, 2017, S. 475). Aufgrund des expliziten Ausweises der beschriebenen Cashflow-Komponenten in der direkten Darstellung, ist die Prognosekraft der direkten, disaggregierten Darstellung des operativen Cashflows erhöht (vgl. Khansalar & Namazi, 2017, S. 475). Im Falle einer indirekten Darstellung des operativen Cashflows müssen die fehlenden Cashflow-Komponenten hingegen indirekt hergeleitet werden. Dabei zeigen Krishnan und Largay III (2000, S. 234), dass eine solche indirekte Herleitung der Cashflow-Komponenten nur mit signifikantem Fehler erfolgen kann. Etwaige Cashflow-Prognosen auf Basis indirekt hergeleiteter Cashflow-Komponenten werden somit einer erhöhten Ungenauigkeit unterliegen (vgl. Orpurt & Zang,

2009, S. 918).

Zusammenfassend ist festzustellen, dass die Prognosekraft der direkten Darstellung des Cashflows aus betrieblicher Tätigkeit durch den expliziten Ausweis einzelner Cashflow-Komponenten erhöht wird. Da diese Cashflow-Komponenten zugleich aus der indirekten Darstellung nicht ohne Fehler herleitbar sind, zeigt sich anhand der betrachteten Studien, dass die direkte Darstellung des Cashflows aus betrieblicher Tätigkeit hinsichtlich ihrer Prognosekraft und somit ihrer Aussagekraft, der indirekten Darstellung überlegen ist.

6. Fazit und Ausblick

Die praktische Bedeutung der Kapitalflussrechnung für den unternehmensexternen Adressaten des Jahresabschlusses zeigt sich in der vielseitigen Nutzung der Kapitalflussrechnung: Neben ihrer unstrittig hohen Aussagekraft als Indikator der Finanzkraft, wird die Kapitalflussrechnung insbesondere auch für den Zweck der Unternehmensbewertung betrachtet. Adressaten, die darüber hinaus vermeintlich verzerrungsfreie Cashflow-Größen als Grundlage zur Plausibilisierung des Periodenergebnisses heranziehen, sollten allerdings bedenken, dass die Cashflow-Größen der Kapitalflussrechnung nach IAS 7 sehr wohl bilanzpolitischen Verzerrungen unterliegen können, welche die Vergleichbarkeit der Kapitalflussrechnung signifikant beschränken. Da die vermeintlich geringe bilanzpolitische Beeinflussbarkeit häufig ein zentrales Argument für die Aussagekraft der Kapitalflussrechnung darstellt, ist diese Erkenntnis von besonderer Bedeutung und impliziert, dass die Aussagekraft der Kapitalflussrechnung in dieser Hinsicht überschätzt werden könnte. Neben dem Ermessensspielraum, der dem Unternehmen bei der Abgrenzung des Finanzmittelfonds gewährt wird, ist insbesondere das Wahlrecht zwischen direkter und indirekter Darstellung des Cashflows aus betrieblicher Tätigkeit, als wichtiger bilanzpolitischer Gestaltungsspielraum hervorzuheben: Wie in Kapitel 5.4 gezeigt, reduziert ein Unternehmen durch die Ausübung des Wahlrechtes zur indirekten Darstellung die Aussagekraft seiner Kapitalflussrechnung signifikant. Durch die Empfehlung für die direkte Darstellung des Cashflows aus betrieblicher Tätigkeit im IAS 7.19 zeigt sich zwar, dass dieses Problem durchaus erkannt wurde. Da in der Rechnungslegungspraxis allerdings die indirekte Darstellung deutlich überwiegt, besteht hier möglicherweise Handlungsbedarf für eine restriktive Vorgabe der direkten Darstellung des Cashflows aus betrieblicher Tätigkeit. Tatsächlich gab es bereits entsprechende Bemühungen, die direkte Darstellung des Cashflows aus betrieblicher Tätigkeit im Rahmen der Kapitalflussrechnung nach IAS 7 vorzuschreiben: Bereits im Jahr 2010 hat das International Accounting Standards Board (IASB) einen Staff Draft veröffentlicht, der unter anderem eine Verpflichtung zur Veröffentlichung einer direkten Darstellung, ergänzend zur indirekten Darstellung des Cashflows aus betrieblicher Tätigkeit, vorsah (vgl. Haller et al., 2014, S. 303f.). Dieser Vorstoß des IASB wurde jedoch letztlich nicht weiter verfolgt (vgl. Bödecker, Busack & Teuteberg, 2016, S. 406f.). Aufgrund der erhöhten Aussagekraft, die eine direkte Darstellung des Cashflows aus betrieblicher Tätigkeit dem Adressaten des Jahresabschlusses bietet, wäre eine erneute Prüfung der ursprünglich geplanten Änderung wünschenswert.

Während die bilanzpolitischen Spielräume, die der IAS 7 in seiner jetzigen Fassung gewährt die Aussagekraft der Kapitalflussrechnung reduzieren, ergibt sich aus der vergleichenden Betrachtung der Aussagekraft von Cashflow-Größe und Periodenergebnis ein geteiltes Bild. Auf der einen Seite legen die betrachteten Wertrelevanz-Studien aus Kapitel 5.2 nahe, dass gerade die Periodenabgrenzungen als wesentliche Differenz zwischen Periodenergebnis und Cashflow-Größe tatsächlich die Aussagekraft des Periodenergebnisses gegenüber der Aussagekraft der Cashflow-Größe erhöhen. Entsprechend folgt aus diesen Studien, dass das Periodenergebnis aufgrund der Berücksichtigung von Periodenabgrenzungen konzeptionell besser geeignet ist, um den Unternehmenserfolg aussagekräftig abzubilden. Auch eine relativ höhere Prognosekraft des Periodenergebnisses zeigt sich in einer Reihe von Studien. Auf der anderen Seite existieren jedoch auch Studien, die den Cashflow-Größen eine höhere Prognosekraft und somit Aussagekraft zuordnen. Diese widersprüchlichen Ergebnisse beruhen auf Abweichungen im Studienaufbau, insbesondere in Bezug auf die Definition der Cashflow-Größe. Aufgrund dieser Abweichungen in Studienaufbau und -ergebnis lässt sich die Frage, welche der beiden Rechnungslegungsgrößen besser für die Cashflow-Prognose geeignet ist, nicht auf Basis der vorliegenden Literatur beantworten. Hier sind weitere Untersuchungen notwendig, die insbesondere Unterschiede im Studienaufbau vorheriger Studien stärker in den Fokus rücken.

Literatur

Ballwieser, W. & Hachmeister, D. (2016). *Unternehmensbewertung: Prozess, Methoden und Probleme* (Bd. 5). Stuttgart, Schäffer-Poeschel Verlag.

Barth, M., Cram, D. & Nelson, K. (2001). Accruals and the Prediction of Future Cash Flows. *The Accounting Review*, *76* (1), 27–58.

Bödecker, A., Busack, H. & Teuteberg, T. (2016). Aktuelle Entwicklungen und ausgewählte Anwendungsfälle zur Zahlungsstromklassifizierung in der Konzernkapitalflussrechnung nach IAS 7 (Teil 1). *Zeitschrift für internationale und kapitalmarktorientierte Rechnungslegung*, *16* (9), 404–409.

Bösser, J., Pilhofer, J. & Lessel, M. (2013). Kapitalflussrechnung nach IAS 7 in der Unternehmenspraxis: Empirische Analysen ausgewählter Aktienindizes. *Praxis der internationalen Rechnungslegung*, *9* (12), 359–366.

Broome, O. (2004). Statement of Cash Flows: Time for Change! *Financial Analysts Journal*, *60* (2), 16–22.

Brösel, G. (2017). *Bilanzanalyse: Unternehmensbeurteilung auf der Basis von HGB- und IFRS-Abschlüssen*. Berlin, Erich Schmidt Verlag.

CFA Institute Centre for Financial Market Integrity. (2007). *A Comprehensive Business Reporting Model: Financial Reporting for Investors*. https://www.cfainstitute.org/-/media/documents/article/position-paper/comprehensive-business-reporting-model.ashx. (Abfrage: 18.10.2021)

Charitou, A., Clubb, C. & Andreou, A. (2000). The Value Relevance of Earnings and Cash Flows: Empirical Evidence for Japan. *Journal of International Financial Management and Accounting*, *11* (1), 1–22.

Cheng, C. & Hollie, D. (2008). Do core and non-core cash flows from operations persist differentially in predicting future cash flows? *Review of Quantitative Finance and Accounting*, *31* (1), 29–53.

Cheng, C., Liu, C.-S. & Schaefer, T. (1996). Earnings Permanence and the Incremental Information Content of Cash Flows from Operations. *Journal of Accounting Research*, *34* (1), 173–181.

Coenenberg, A., Haller, A. & Schultze, W. (2021). *Jahresabschluss und Jahresabschlussanalyse: Betriebswirtschaftliche, handelsrechtliche, steuerrechtliche und internationale Grundlagen - HGB, IAS/IFRS, US-GAAP, DRS* (Bd. 26). Stuttgart, Schäffer-Poeschel Verlag.

Coenenberg, A. & Meyer, M. (2003). Kapitalflussrechnung als Objekt der Bilanzpolitik. In M. Henke, U. Jung, N. Schönbrunn, H. Siebert & P. Wollmert (Hrsg.), *Wirtschaftsprüfung und Unternehmensüberwachung. Festschrift für Prof. Dr. Dr. h.c. Wolfgang Lück* (S. 335–383). Düsseldorf, IDW Verlag.

Dechow, P. (1994). Accounting earnings and cash flows as measures of firm performance: The role of accounting accruals. *Journal of Accounting and Economics*, *18* (1), 3–42.

Dechow, P., Kothari, S. & Watts, R. (1998). The relation between earnings and cash flows. *Journal of Accounting and Economics*, *25* (2), 133–168.

DeFond, M. & Hung, M. (2003). An empirical analysis of analysts' cash flow forecasts. *Journal of Accounting and Economics*, *35* (1), 73–100.

Eiselt, A. & Müller, S. (2014). *Kapitalflussrechnung nach IFRS und DRS 21: Darstellung und Analyse von Cashflows und Zahlungsmitteln*. Berlin, Erich Schmidt Verlag.

Farshadfar, S. & Monem, R. (2013). The usefulness of operating cash flow and accrual components in improving the predictive ability of earnings: a re-examination and extension. *Accounting and Finance*, *53* (4), 1061–1082.

FASB. (1978). *Statement of Financial Accounting Concepts No. 1: Objectives of Financial Reporting by Business Enterprises*. https://www.fasb.org/jsp/FASB/Document_C/DocumentPage?cid=1218220132541&acceptedDisclaimer=true. (Abfrage: 18.10.2021.)

Finger, C. (1994). The Ability of Earnings to Predict Future Earnings and Cash Flow. *Journal of Accounting Research*, *32* (2), 210–223.

Gebhardt, G. & Mansch, H. (2012). Praxis der Aufstellung und Nutzung von Kapitalflussrechnungen deutscher Industrieunternehmen. *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung – Sonderhefte*, *41* (2).

Greenberg, R., Johnson, G. & Ramesh, K. (1986). Earnings versus Cash Flow as a Predictor of Future Cash Flow Measures. *Journal of Accounting, Auditing & Finance*, *1* (4), 266–277.

Haller, A., Groß, T. & Rauscher, A. (2014). Notwendigkeit und Ansatzpunkte für eine Reform des Statement of Cashflows nach IAS 7: Empirische Evidenz aus der deutschen Rechnungslegungspraxis. *Zeitschrift für internationale und kapitalmarktorientierte Rechnungslegung*, *14* (6), 298–306.

Hitz, J.-M. & Teuteberg, T. (2013). Verpflichtende und freiwillige Cashflow-Berichterstattung: Empirischer Befund für den deutschen Kapitalmarkt. *Zeitschrift für internationale und kapitalmarktorientierte Rechnungslegung*, *13* (1), 33–42.

Hribar, P. & Collins, D. (2002). Errors in Estimating Accruals: Implications for Empirical Research. *Journal of Accounting Research*, *40* (1), 105–134.

Hüttche, T. & Int-Veen, T. (2008). Praxis der Kapitalflussrechnung: Empirische Untersuchung der Bilanzpolitik - Konsequenzen für die Analyse. *Zeitschrift für internationale Rechnungslegung*, *3* (7/8), 347–354.

IAS. (2016). *IAS 7: International Accounting Standard 7 "Statements of Cash Flows" (1992, revised 2016)*. (zuletzt geändert im Rahmen der "Disclosure Initiative", London)

Jones, S. & Widjaja, L. (1998). The Decision Relevance of Cash-Flow Information: A Note. *Abacus*, *34* (2), 204–219.

Kaiser, D. & Bierwulf, L. (2011). Analyse der Cash Flow Statements der DAX-Unternehmen. *Corporate Finance biz*, *2* (1), 3–12.

Khansalar, E. & Namazi, M. (2017). Cash flow disaggregation and prediction of cash flow. *Journal of Applied Accounting Research*, *18* (4), 464–479.

Krishnan, G. & Largay III, J. (2000). The Predictive Ability of Direct Method Cash Flow Information. *Journal of Business Finance & Accounting*, *27* (1/2), 215–245.

Kühnberger, M. & Thurmann, P. (2015). Zur Bedeutung von Gewinnen und Cashflows in der empirischen Rechnungslegungsforschung. *Zeitschrift für internationale und kapitalmarktorientierte Rechnungslegung*, *15* (1), 30–43.

Lachnit, L. (2004). *Bilanzanalyse: Grundlagen - Einzel- und Konzernabschlüsse - internationale Abschlüsse - Unternehmensbeispiele*. Wiesbaden, Gabler Verlag.

Lachnit, L. & Müller, S. (2017). *Bilanzanalyse: Grundlagen – Einzel- und Konzernabschlüsse – HGB- und IFRS-Abschlüsse – Unternehmensbeispiele* (Bd. 2). Wiesbaden, Springer Gabler Verlag.

Liu, J., Nissim, D. & Thomas, J. (2007). Is Cash Flow King in Valuations? *Financial Analysts Journal*, *63* (2), 56–68.

Lüdenbach, N., Hoffmann, W. & Freiberg, J. (2020). *Haufe IFRS-Kommentar*. Freiburg, Haufe Group.

Meyer, M. (2007). *Cashflow-Reporting und Cashflow-Analyse: Konzeption, Normierung, Gestaltungspotenzial und Auswertung von Kapitalflussrechnungen im internationalen Vergleich*. Düsseldorf, IDW Verlag.

Mostafa, W. (2014). The relative information content of cash flows and earnings affected by their extremity: UK evidence. *Managerial Finance*, *40* (7), 646–661.

Müller, S. (2008). Die Kapitalflussrechnung als zentrales Instrument des konvergenten Rechnungswesens. In C.-C. Freidank, S. Müller & I. Wulf (Hrsg.), *Controlling und Rechnungslegung: Aktuelle Entwicklungen in Wissenschaft und Praxis. Festschrift für Prof. Dr. Laurenz Lachnit zum 65. Geburtstag* (S. 285–300). Wiesbaden, Gabler Verlag.

Nallareddy, S., Sethuraman, M. & Venkatachalam, M. (2020). Changes in accrual properties and operating environment: Implications for cash flow predictability. *Journal of Accounting and Economics*, *69* (2/3), 1–23.

Nowak, K. (2003). Marktorientierte Unternehmensbewertung: Discounted Cash Flow, Realoption, Economic Value Added und der Direct Comparison Approach. In H.-J. Böcking & M. Hommel (Hrsg.), *Rechnungswesen und Unternehmensüberwachung* (Bd. 2). Wiesbaden, Deutscher Universitäts-Verlag.

Oehlrich, M. (2020). Ermittlung des Free Cash Flows in der Kapitalflussrechnung nach DRS, IFRS und US-GAAP. *Zeitschrift für Internationale Rechnungslegung*, *15* (7/8), 355–359.

Ordelheide, D. & Leuz, C. (1998). Die Kapitalflußrechnung: Grundlagen, International Accounting Standard No. 7 und Informationsgehalt. *Wirtschaftswissenschaftliches Studium*, *27* (4), 176–183.

Orpurt, S. & Zang, Y. (2009). Do Direct Cash Flow Disclosures Help Predict Future Operating Cash Flows and Earnings? *The Accounting Review*,

84 (3), 893–935.

Pollmann, R. (2012). Struktur und Gliederung der Kapitalflussrechnung nach IAS 7. *Zeitschrift für internationale Rechnungslegung*, *7* (5), 191–198.

Rehkugler, H. & Poddig, T. (1998). *Bilanzanalyse* (Bd. 4). München, R. Oldenbourg Verlag.

Ruprecht, R. & Nösberger, T. (2012). Kapitalflussrechnungen nach IAS 7: Praktische Probleme. *Zeitschrift für Internationale Rechnungslegung*, *7* (3), 109–114.

Schultze, W. & Meyer, M. (2005). Die Kapitalflussrechnung in der Unternehmensbewertung: Bewertungskalkül und Gestaltungshinweise. *Zeitschrift für Planung & Unternehmenssteuerung*, *16* (3), 289–324.

Wulf, I. (2008). Wirkung der IFRS-Bilanzierung auf die Höhe von Jahresergebnis und Cash-flow: Eine Fallstudie unter Berücksichtigung ausgewählter Bilanzierungssachverhalte beim Übergang von HGB auf IFRS. *Zeitschrift für internationale und kapitalmarktorientierte Rechnungslegung*, *8* (10), 649–656.

Junior Management Science 8(1) (2023) 237-269

Junior Management Science

journal homepage: www.jums.academy

Do birds of a feather always flock together? A multidimensional examination of homophily in crowdfunding

Markus Klepsch

Technische Universität München

Abstract

Homophily—the tendency of individuals to associate with similar others—is one of the most persistent findings in research on interpersonal interaction. Literature has recently also studied the impact of homophily in crowdfunding markets (Greenberg & Mollick, 2017). While these results offer valuable insights into the dimension sex, homophily is a multidimensional construct in theory (McPherson, Smith-Lovin, & Cook, 2001). Therefore, I employ homophily theorizing by analyzing to what extent homophily influences backers' choices in crowdfunding regarding five different sociodemographic dimensions. To test my hypotheses, I drew upon a unique dataset of roughly 3.5 million backings made on the crowdfunding platform Indie-gogo. My results suggest that individuals make homophilic choices with respect to age, sex, occupation, and location. While my findings thus confirm that, in large, homophily plays an essential role in individual choices, I extend the literature by showing that it is not always as clear-cut. The more complex the dimension, the more difficult it is to judge. Specifically, I show that in the dimension race, where a plethora of possible combinations leads to a certain degree of complexity, there is no simple 'yes' or 'no' answer to whether homophily influences the decision. Instead, significant differences can be observed depending on the specific race under consideration.

Keywords: Crowdfunding; Indiegogo; Homophily.

1. Introduction

Choice homophily (Greenberg & Mollick, 2015, 2017; McPherson & Smith-Lovin, 1987) is influential in resource acquisition online. In general, the concept of homophily can be described as "like attracted to like" (Greenberg & Mollick, 2017, p. 342) and argues that the basis of attraction between individuals is the similarity between them, where "contact between similar people occurs at a higher rate than among dissimilar people" (McPherson et al., 2001, p. 416). In the past, the concept of homophily has been used to explain a plethora of phenomena in the social sciences ranging from mere personal contact (Wellman, 1996) over friendship (Kandel, 1978; Lazarsfeld & Merton, 1954; Verbrugge, 1977) to the close bond of marriage (Alba & Golden, 1986; Kalmijn, 1998). Essentially, two types of homophily—induced and choice—can be distinguished. While induced homophily describes "structural mechanisms that define the probability that individuals will interact with those with similar characteristics" (Greenberg & Mollick, 2017, p. 342), choice homophily is acting at the individual level. It explains how people choose to interact with similar others due to shared characteristics (Greenberg & Mollick, 2017). The concept of choice homophily has already been used to explain resource acquisition in different settings such as crowdlending (Riggins & Weber, 2017), venture capital investment (Hegde & Tumlinson, 2014), and crowdfunding (Greenberg & Mollick, 2017). However, it should be noted that homophily often was studied regarding only a single characteristic, such as *sex and gender*. For example, a first study in the context of crowdfunding demonstrated how female backers tend to support female entrepreneurs and even more so in industries where women are underrepresented (Greenberg & Mollick, 2017).

However, homophily is a multidimensional construct. It is not only related to sex and gender, but amongst others also to *race and ethnicity*, *age*, and *occupation* (McPherson et al., 2001). Therefore, it is necessary to include other, often neglected dimensions in the analysis to advance research on homophily, especially in online fundraising. Moreover, homophily occurs in the dyadic relationship between the investor and the entrepreneur in the online fundraising context. Especially in crowdfunding, however, research has so far focused primarily on the entrepreneur, neglecting the role

DOI: https://doi.org/10.5282/jums/v8i1pp237-269

of the investors. By studying how investors choose to support entrepreneurs, this paper aims to put the investors at the center of research.

Since homophily is a multidimensional construct in theory, it is of concern to adopt a multifaceted view in the online resource acquisition context to advance the understanding of the investors' decision-making. Moreover, it is simply not clear whether the findings for the characteristic sex are reproducible for other dimensions of homophily, such as race and ethnicity, age, occupation, or location. While one possible outcome of the analysis might be that investors exhibit homophily with regard to other dimensions, different mechanisms may just as well be responsible for the behavior of investors. In this context, the concept of intragroup competition (Deutsch, 1949; Goldman, Stockbauer, & McAuliffe, 1977), occurring due to perceived conflicting goal attainments among members of a group, could provide an alternative explanation for the behavior of investors. This scenario is particularly imaginable in the dimension occupation. Backers who have launched a crowdfunding campaign themselves may not fund projects in the same category as theirs since this may decrease the probability of success of their own campaign. Due to a lack of research to date, it is simply unclear to what extent homophily influences the investors' decision-making in online fundraising. Thus, the research question underlying this paper is as follows: *To what extent does homophily play a role in the choice of crowdfunding investors, and specifically, to what extent do different sociodemographic characteristics influence homophily?*

To answer the research question, I perform quantitative analyses on data gathered from Indiegogo, one of the largest crowdfunding platforms worldwide. For this purpose, I constructed a dataset of roughly 3.5 million backings made by more than one million backers on Indiegogo. Using a self-built web crawler, I collected various data points such as the name, location, profile image, and social media links of the backers and entrepreneurs. Thereby, my quantitative analyses focus on studying homophily effects in the five sociodemographic dimensions age, sex, race, occupation, and location. The different research hypotheses guiding my analyses can be summarized as follows: Investors are more likely to support those entrepreneurs to whom they are similar.

This study makes four contributions. First, I contribute to the broad literature on homophily. While scholars often studied homophily regarding single dimensions, for example, in a study demonstrating how women tend to support women in crowdfunding (Greenberg & Mollick, 2017), I extend the frame of analyses and individually consider five different sociodemographic dimensions of homophily. Thus, this study is one of the first to give a broader overview of homophily in different dimensions to do more justice to the complexity of this theoretical construct. In this context, I find that, as previously demonstrated, individuals tend to make homophilous choices regarding the sociodemographic dimensions age, sex, occupation, and location. In contrast, homophily in the more complex dimension race is not as clear-cut. Instead, I find significant differences depending on the race considered, where my findings suggest that primarily black and Indian individuals make homophilous choices.

Second, I contribute to the crowdfunding literature by moving the investors into the spotlight of crowdfunding research. Scholars have previously called to put the investors in the spotlight, as so far, research has mainly focused on the characteristics of the campaigns and the entrepreneurs as drivers of success, while the backers have been neglected in the crowdfunding context (Bretschneider & Leimeister, 2017; Kuppuswamy & Bayus, 2018). Analyzing the decision-making behavior of the investors also offers valuable practical insights for potential campaign creators. In this context, my findings suggest that black and Indian entrepreneurs are discriminated against by every but their race.

Third, most studies in crowdfunding obtained their data from the platform Kickstarter (exemplary see Greenberg and Mollick (2017)). My study empirically contributes by collecting the data on the backers' investment behavior from the crowdfunding platform Indiegogo.

Fourth, this study contributes by replicating the findings observed on other platforms for resource acquisition online on the crowdfunding platform Indiegogo. While, for example, a study on the lending-based microfinance platform Kiva showed that lenders tend to support fundraisers of the same sex and projects within an industry similar to the one the lenders are active in (Riggins & Weber, 2017), it is not clear whether these results are projectable to other platforms like Indiegogo. That is because scholars have argued that platforms differ, and thus "patterns observed on one platform cannot be assumed to generalize to other platforms" (Dushnitsky & Fitza, 2018, p. 1). My results, however, help to replicate and validate previous findings from other platforms for online fundraising and thus contribute by providing an additional perspective to the question of platform generalizability.

2. Theory and Hypotheses

The following chapter provides a review of the theories motivating this study. This chapter starts with a review of the general theory of homophily (see chapter 2.1), followed by an introduction to homophily in the context of crowdfunding (see chapter 2.2). Finally, this chapter ends with a theoretical explanation, including the hypothesis development, for each of the five sociodemographic dimensions of homophily examined in this paper (see chapter 2.3).

2.1. Homophily

The term *homophily* was initially coined by Lazarsfeld and Merton (1954) to describe the tendency of individuals to associate with similar others and to provide a clear distinction from the complementary term *heterophily*, which denotes a tendency for people who differ in certain respects to interact. The observation itself that people bond with others who are similar to them is rather old. Already Aristotle stated in his Rhetoric and Nichomachean Ethics that people "love

those who are like themselves" (Aristotle, 1934, p. 1371), or Plato explained in Phaedrus that "similarity begets friendship" (Plato, 1968, p. 837).

Today, the phrase "birds of a feather flock together" (Lazarsfeld & Merton, 1954, p. 37) is commonly used to summarize the concept of homophily. In general, this theoretical construct describes a psychological process suggesting that similarity promotes connection, whereby bonds between similar people occur more frequently than between dissimilar people (McPherson et al., 2001). However, it is important to stress that homophily does not only occur in one dimension, such as race and ethnicity. Instead, it is a multidimensional construct, where race and ethnicity, age, sex and gender, religion, education, and occupation are among the main dimensions where similarity fosters interaction (McPherson et al., 2001).

The tendency of individuals to associate with others based on similar characteristics is one of the most thoroughly studied concepts in the social sciences (Greenberg & Mollick, 2017). In the past, homophily has been used to explain a plethora of phenomena in this field, such as personal contact (Wellman, 1996), friendship (Kandel, 1978; Lazarsfeld & Merton, 1954; Verbrugge, 1977), and marriage (Alba & Golden, 1986; Kalmijn, 1998). The overarching finding of all studies is that, for example, friends, colleagues, and individuals in romantic relationships are more similar to each other on various dimensions than randomly selected members of a population (Kossinets & Watts, 2009). In this context, early studies related to homophily and network formation were conducted among small social groups, where ties between group members could easily be observed by ethnographers, such as people meeting at a cafeteria, behavior among school children, or small urban neighborhoods (McPherson et al., 2001). Thereby, studies have demonstrated homophily by psychological traits such as aspirations, attitudes, and intelligence, suggesting that a community of values plays a vital role in the formation of friendships (Richardson, 1940). Other studies in the early phase were able to demonstrate significant homophily by demographic factors like sex, race and ethnicity, education, and age (exemplary see Bott (1928); Loomis (1946)).

With advances in research methods, scholars conducted large-scale studies of network formation and homophily beginning in the 1970s (McPherson et al., 2001). By taking advantage of more advanced research methods, studies on adolescent friendships were able to show that social status significantly influences friendships, as adults tend to choose friends with an equal social status (Verbrugge, 1977). Furthermore, studies on the formation and dissolution of adult friendships showed that adults choose friends with similar attitudes and that an incongruence of attitudes and behaviors can lead to friendship breakup (Kandel, 1978). In addition, studies that uncovered the mechanisms of the closest bond of marriage were able to show that weddings in the United States of America are more likely to happen between individuals with ethnically related backgrounds (Alba & Golden, 1986). Another study worth mentioning examined the influence of local ties, i.e., the neighborhood in which people live, on personal networks (Wellman, 1996). This study provided significant results showing the importance of local ties in personal networks, emphasizing the importance of the local dimension in interpersonal contact (Wellman, 1996). While the studies presented above, representing only a tiny portion of the vast literature on homophily, may differ in their contexts and methods, the primary finding that runs like a thread through all of the studies is that "people's personal networks are homogeneous with regard to many sociodemographic, behavioral, and intrapersonal characteristics" (McPherson et al., 2001, p. 415).

From those studies, two important aspects should be highlighted. First, most studies on homophily refer to the level of individual interaction, which describes the dyadic relationship between two individuals, such as the study on ethnic marriage (Alba & Golden, 1986). Besides the dyadic relationship between two individuals, the community level, which assumes that peer groups play an essential role in people's behavior, forms the other main tradition of research on homophily (McPherson et al., 2001). Second, the majority of the studies consider only one of the dimensions of homophily (Block & Grund, 2014), such as the influence of race and ethnicity on marriage (Alba & Golden, 1986). However, some scholars also emphasize the multidimensionality in network formation and social structure by measuring their development on several characteristics. In this regard, a study of social structure formation was one of the first to include multiple factors such as education, sex, and many more in the analysis (Blau, 1977).

Furthermore, two types of homophily—induced homophily and choice homophily—need to be distinguished. These different types of homophily describe the two possible mechanisms in which the similarity of related individuals may be grounded. First, induced homophily is caused by the group composition (Kossinets & Watts, 2009; McPherson & Smith-Lovin, 1987) and describes structural mechanisms defining the probability of interaction with similar others (Greenberg & Mollick, 2017). Thus, even in the case of random choices within the group, the social ties would create similarity between individuals (Greenberg & Mollick, 2017), suggesting that the group composition in the system dictates the possibility for association (McPherson & Smith-Lovin, 1987). On the other hand, choice homophily—the focus of this paper—operates at the individual level, is driven by shared characteristics, and explains how people choose to support and interact with similar others due to personal preferences (Greenberg & Mollick, 2017; Kossinets & Watts, 2009). Thus, in the case of pure choice homophily, the group's composition would not affect similarity between connected individuals, but the interpersonal interaction would be based solely on preferences for the dyadic similarity between the connected individuals (Greenberg & Mollick, 2017; McPherson & Smith-Lovin, 1987). Specifically, interaction is based on individual choices, with interpersonal similarity being the cause of association with someone similar to oneself (Greenberg & Mollick, 2015).

One possible reason for choice homophily presented in the literature is that being akin to someone—e.g., sharing the same cultural background, being the same age, or speaking the same language—can lead to trust and solidarity between individuals (Block & Grund, 2014). In short, choice homophily is "the individual-level propensity to choose similar others" (McPherson & Smith-Lovin, 1987, p. 371). However, it should be noted that historically it has been challenging to separate choice homophily from induced homophily (Greenberg & Mollick, 2015). That is because the different forms often do not occur in isolation but rather in some sort of interaction (Kossinets & Watts, 2009; McPherson & Smith-Lovin, 1987). Thereby, "groups are arenas for tie formation" (McPherson & Smith-Lovin, 1987, p. 373), where a group can be as small as a school class or large like platforms for online venture funding. Thus, opportunity structures for interactions are influenced by the size, composition, and structure of the group (McPherson, 1982, 1983; McPherson & Smith-Lovin, 1982, 1986), describing how induced homophily shapes choice homophily as structural barriers set the context on the options individuals can choose from (Greenberg & Mollick, 2017).

Over time, the concept of homophily has been applied not only in sociological research but also in other fields, such as business and economics. For example, a study examining the formation of inter-organizational alliances found that various rules of affiliation such as homophily, accumulative advantage, and follow-the-trend influence network evolution (Powell, White, Koput, & Owen-Smith, 2005). Those results are consistent with the findings of a study on structural homophily and firm partnering choices, which found that homophily can frequently be used to explain organizational alliances (Ahuja, Polidoro Jr, & Mitchell, 2009). Interestingly, a study focusing on interactions between employees within an organization found that employees exhibit significant homophily in communicating with other employees, especially in large groups that offer more choice of interaction partners than smaller groups (Kleinbaum, Stuart, & Tushman, 2013). This finding already points to the relevance of choice homophily, especially in large groups. Therefore, the concept of choice homophily is particularly relevant in the context of online venture funding, especially crowdfunding, as it involves large groups, and studies in this context aim to explain how investors choose to fund specific projects and entrepreneurs. For example, in this setting, a study examining the influence of social proximity on business partnerships found that venture capitalists tend to choose start-ups with co-ethnic executives for their investment, describing an ethnic proximity between the venture capital firm and the start-up (Hegde & Tumlinson, 2014). Similarly, an initial study of homophily in the context of crowdfunding—which will be mentioned only briefly here and discussed in more detail in the following chapter—found that females tend to support female entrepreneurs (Greenberg & Mollick, 2017). The above studies show that the concept of choice homophily has gained interest in the business and economics contexts, where choice homophily has been used to explain various settings of entrepreneurial funding. Thereby, recent interest has also turned to crowdfunding markets, as they provide a near-optimal framework to focus almost solely on choice homophily (Greenberg & Mollick, 2017).

2.2. Homophily in Crowdfunding

As explained in chapter 2.1, homophily has already been applied in various fields of literature, ranging from sociology to business and economics. Focusing almost purely on choice homophily, an initial study in the context of crowdfunding demonstrated that women tend to support other women, and even more so in industries where women are typically underrepresented (Greenberg & Mollick, 2017). In this context, Greenberg and Mollick (2017) further divide choice homophily into two more detailed mechanisms in order to differentiate between homophily caused by mere dyadic similarity, i.e., interpersonal choice homophily, and homophily caused by "perceptions of shared structural barriers stemming from a common group-level social identity and an underlying desire to help overcome them" (p. 342), i.e., activist choice homophily.

The goal of the above study was to choose a context in which homophily has historically been viewed as a cause of inequality, namely the funding of female entrepreneurs (Greenberg & Mollick, 2017). Thereby, past studies pointed to the fact that women receive less venture capital than men. For example, women-led firms receive only a small percentage of venture capital investments (Greene, Brush, Hart, & Saparito, 2001; Harrison & Mason, 2007). A possible reason for this gender gap is the inequality in the traditional funding network itself, where less than 15 percent of venture capitalists are female (Harrison & Mason, 2007; Stuart & Sorenson, 2007), having scholars suggest that homophily is one of the reasons for the gender imbalance in venture funding (Becker-Blease & Sohl, 2007; Stuart & Sorenson, 2007). However, connecting to the overarching theory of homophily, it is worth noting that an inequality exists not only in the funding of female entrepreneurs but also in the broader context of resource acquisition concerning other sociodemographic characteristics such as race and ethnicity. In this context, studies have shown that inequality also exists in this regard, as African American men are less likely than white entrepreneurs to obtain funding (Jenq, Pan, & Theseira, 2015; Younkin & Kuppuswamy, 2018). This observation may be due, at least to some degree, because most early-stage investors are white men (Sohl, 2015). To study such biases and to focus merely on the choices made by investors, that is, on the concept of choice homophily, crowdfunding provides a more appropriate context than traditional means of funding such as venture capital firms (Greenberg & Mollick, 2017). Thus, the crowdfunding context offers a solution to the long-standing challenge of separating induced and choice homophily (Greenberg & Mollick, 2015, 2017).

The term crowdfunding was first coined in the late 2000s and "describes a new institutional form which utilizes digital platforms to originate and aggregate funding" (Dushnitsky & Fitza, 2018, p. 1). While early-stage funding and

resource acquisition have always been crucial for commercializing innovative ideas (Cosh, Cumming, & Hughes, 2009; Kortum & Lerner, 2000), deciding which new ideas to commercialize has long been tremendously undemocratic, with small groups of experts, typically white male venture capitalists, deciding which new idea to support (Mollick & Robb, 2016). The lack of diversity in traditional venture funding and the resulting barriers in access to financial resources (Mollick & Robb, 2016) have led to an increasing number of entrepreneurs using a relatively new form of financing that makes use of large online communities, known as *crowdfunding* (Kuppuswamy & Bayus, 2018; Mollick, 2014). In a prevalent definition of the term, crowdfunding "refers to the efforts by entrepreneurial individuals and groups (...) to fund their ventures by drawing on relatively small contributions from a relatively large number of individuals using the internet, without standard financial intermediaries" (Mollick, 2014, p. 2). In this context, the relatively large number of individuals is also commonly referred to as *the crowd* (Belleflamme, Lambert, & Schwienbacher, 2014). Appealing to a large community of investors from the general public, crowdfunding "involves an open call, mostly through the Internet, for the provision of financial resources either in the form of donation or in exchange for the product or some form of reward to support initiatives" (Belleflamme et al., 2014, p. 4). Ultimately, the goal of crowdfunding platforms is to connect entrepreneurs with investors who are willing to support new ideas (Bretschneider & Leimeister, 2017). Thereby, the goal is to democratize the funding process by serving as a vehicle for greater participation by investors and entrepreneurs who are typically underrepresented in these markets, such as women and all sorts of minorities (Mollick & Robb, 2016). While crowdfunding can currently still be considered a niche phenomenon in total funding volume, it is gaining importance in many countries (Belleflamme, Omrani, & Peitz, 2015). The crowdfunding market has grown from US$880 million in 2010 to US$34.4 billion in 2015 (Massolution, 2015) and is expected to grow to US$96 billion by 2025 (World Bank, 2013). Although these figures are comparatively small compared to the trillions of dollars invested overall, they show that crowdfunding is a rapidly growing market (Belleflamme et al., 2015).

As can be inferred from the brief description above, the crowdfunding context is particularly well suited to focus on the concept of choice homophily. By reducing social constraints and search costs, crowdfunding allows campaign creators to access potential backers with similar sociodemographic characteristics and allows individual investors to choose to fund ideas they support (Greenberg & Mollick, 2017). Thus, choosing the crowdfunding context can lower induced homophily in two main ways and allows to focus almost entirely on choice homophily. First, it removes the traditional gatekeepers of venture funding, such as venture capitalists, and at the same time increases the set of potential investors by offering everyone the opportunity to fund a campaign (Greenberg & Mollick, 2017). As shown above, traditional gatekeepers to start-up funding lack diversity with respect to various sociodemographic characteristics, such as sex or race. Thus, since the lack of diversity limits the possibility for connections between investors and entrepreneurs, traditional funding includes high levels of induced homophily, disfavoring minorities that have historically been disadvantaged in the venture funding process. However, crowdfunding potentially democratizes resource acquisition by removing traditional gatekeepers and providing investment opportunities to a possibly unlimited community of investors from the general public. It increases diversity among investors with respect do various sociodemographic characteristics, ultimately reducing induced homophily.

Second, using digital platforms as a means of funding makes it possible to reach a larger group of like-minded individuals than is typically the case with socially or geographically constrained searches (Greenberg & Mollick, 2017). In line with McPherson and Smith-Lovin (1987), arguing that given a certain amount of diversity, larger groups allow for more homophilic ties, crowdfunding, with its inherently large and diversified community, presents the perfect setting to focus on individual choices. In this setting with a large set of possible choices, it is thus possible to properly study choice homophily because if individuals "make homophilous choices, they produce similar pairings even in cases of high group diversity" (McPherson & Smith-Lovin, 1987, p. 372).

2.3. Sociodemographic Dimensions of Homophily

The following chapter aims to theoretically motivate and explain the five sociodemographic dimensions of homophily studied in this paper. I provide a theoretical explanation regarding the backers' possible behavior with respect to the dyadic relationship with the entrepreneur for each dimension analyzed—i.e., *age*, *sex*, *race*, *occupation*, and *location*—including the respective hypotheses guiding the analyses. In line with the goal of this paper, which is to examine how investors choose to support entrepreneurs in online resource acquisition, the following chapter focuses on choice homophily. Thus, this chapter provides theoretical explanations for situations where choice is warranted, describing why investors might show a preference to invest in entrepreneurs to whom they are similar.

2.3.1. Age

Apart from generation-linking ties such as relationships with children or parents, age homophily has been proven to be quite strong across general social relationships (Smith, McPherson, & Smith-Lovin, 2014). However, typically a large part of age homophilic networks is induced by institutional settings and life-course patterns such as workplaces, schools, or sports clubs (Kalmijn & Vermunt, 2005). These findings are consistent with a study by Feld (1982), who argues that the environments in which people engage tend to be homogeneous, indicating that the group's composition implies the extent of age homophily in a network.

Apart from age homophilic networks induced by structural and institutional environments, scholars have also presented mechanisms that explain the emergence of age homophilic networks due to individual choice. For example, a study in the social network context was able to transfer the findings of age homophily from the offline world to the online world, specifically to the social network Myspace (Thelwall, 2009). This finding is interesting because although the internet is known for increasing the diversity of opinions, information, and sociodemographic characteristics of the pool of an individual's potential contacts, the study shows that people still choose to connect with similar others (Thelwall, 2009). Thus, the results of this study suggest that age homophily is not only induced but also a result of choice.

In addition, scholars have argued that even minor age differences can lead to significant differences in individuals' interests (Smith et al., 2014), suggesting that individuals choose to interact with other individuals who are similar in age based on shared interests. This finding is consistent with studies on age differences and changes in activities, demonstrating a shift in activities as people age due to changes in individual preferences, abilities, and constraints (Verbrugge, Gruber-Baldini, & Fozard, 1996). This idea also seems plausible in the context of online resource acquisition. It is conceivable that individuals of a similar age share similar interests. Therefore, an investor might be particularly attracted to the product or service of an entrepreneur of similar age because they both share the same interests, possibly reflected in the product or service.

In addition, the theories of social identity (Tajfel, 1974) and self-categorization (Turner, Hogg, Oakes, Reicher, & Wetherell, 1987) provide a possible explanation for the investors' age-homophilic behavior. The process of classifying individuals into similar others (in-group members) and dissimilar others (out-group members) is a basic psychological process rooted in the desire to increase self-esteem. Thereby individuals tend to exhibit a bias in favor of members of their group (Avery, McKay, & Wilson, 2007). This categorization of others and oneself and associating values with the various social categories (Tajfel, 1974; Turner & Oakes, 1986) is related to the goal of individuals to maintain persistent identities (Steele, 1988), which explains why individuals tend to positively value similar individuals over dissimilar individuals (Goldberg, 2005). Please note that the above paragraph considers the theories therein in light of the sociodemographic dimension age. However, these theories can, in principle, explain the behavior in terms of a variety of different demographic attributes defining individuals, as according to those theories, individuals classify others and themselves based on various dimensions such as age, gender, or race (Avery et al., 2007). Therefore, I will also refer to what is presented here in the following chapters to explain possible outcomes in other dimensions.

Consistent with this line of reasoning, the framework of relational demography (Tsui & Gutek, 1999; Tsui & O'reilly, 1989) argues that individuals act more favorably in environments with a higher number of in-group members in terms of demographic characteristics like age. Thus, similarity in demographic characteristics increases perceived similarity in experiences and values, thus promoting coherence and identification among in-group members (Mehra, Kilduff, & Brass, 1998). Additionally, the similarity-attraction paradigm (Byrne, 1971), closely related to the social identity theory, should also briefly be mentioned here. This paradigm argues that similar individuals are attracted and experience positive outcomes in interaction and association (Goldberg, 2005).

Similarly, it seems plausible that investors in online resource acquisition tend to support entrepreneurs of similar age, both because of shared interests, possibly leading to a high commonality of interest in the funded project, and because of self-categorization and social identity. Therefore, I propose:

> *Hypothesis 1 (H1): Contributions to entrepreneurs with a similar age as the investor occur more frequently than we would expect if investors chose projects for contribution at random (baseline).*

2.3.2. Sex and Gender

As already mentioned, an initial study on homophily in the context of crowdfunding was able to show that women choose to support each other (Greenberg & Mollick, 2017). In this context, a vital driver of the favored behavior regarding similar individuals is the basic principle of choice homophily, which states that individuals, on average, have a positive affection for common characteristics (Huston & Levinger, 1978; Ingram & Morris, 2007; Lazarsfeld & Merton, 1954; Verbrugge, 1977). Scholars propose different drivers of choice homophily in this context, where these drivers include, among others, trust and ease of communication (Kossinets & Watts, 2009; Wimmer & Lewis, 2010). However, please note that the principle of choice homophily—attraction due to shared characteristics—can be used to explain the behavior in terms of any sociodemographic characteristic shared by two individuals (Wimmer & Lewis, 2010). Thus, the principle of choice homophily represents a central pillar for explaining the possible behavior of investors, which I will refer to throughout chapter 2.3.

The preference for shared characteristics, commonly used to explain the emergence of same-sex ties, is consistent with what is presented for the theories of social identity and self-categorization (see chapter 2.3.1). Those theories describe how individuals classify themselves and others and prefer ties to in-group members in part to increase self-esteem (Avery et al., 2007). In the context of increasing self-esteem, another argument seems plausible to describe the tendency of individuals to support others of the same gender in the process of resource acquisition, which is that individuals "seek information affirming identification with their in-groups" (Avery et al., 2007, p. 1543). This argument suggests that individuals do not blindly look for similarity, such as in terms of gender, as a basis for engagement but instead look for identity-affirming similarity (Avery et al., 2007). In short, this means

that the bias only occurs when the members of the in-group that are potentially interacted with help maintain and create a positive impression of that group (Avery et al., 2007). This argumentation is particularly applicable to resource acquisition online because, in this context, investors choose to support entrepreneurs, where the entrepreneur is generally positively portrayed as an individual who aims to create wealth and foster innovation (Carland, Hoy, & Carland, 1988).

In addition, the same-gender bias effect (Mobley, 1982) could provide a possible theoretical explanation for the behavior of investors in choosing whom to support or not support in online venture funding. This principle has shown that, for example, in an organizational setting, raters rate members of their gender subgroup higher than members of other subgroups (Mobley, 1982). Similarly, it is conceivable that not only in organizational settings but also in online resource acquisition, raters (i.e., the investors) tend to value members of their gender subgroup higher than members of other subgroups and therefore show a tendency to support entrepreneurs with whom they share the same gender.

In line with the results expected from the above argumentation, several studies in the context of crowdfunding were also able to demonstrate that investors show a tendency to fund entrepreneurs of the same gender (Greenberg & Mollick, 2015, 2017; Groza, Groza, & Barral, 2020; Riggins & Weber, 2017; Vismara, Benaroio, & Carne, 2017). Thus, in line with the theories presented above as well as findings from previous studies, I propose:

> *Hypothesis 2 (H2): Contributions to entrepreneurs of the same sex as the investor occur more frequently than we would expect if investors chose projects for contribution at random (baseline).*

2.3.3. Race and Ethnicity

Race and ethnicity undeniably present the most significant divide in social networks (McPherson et al., 2001). Thereby, research has demonstrated significant racial homophily in a variety of relationships ranging from school friendships (Shrum, Cheek Jr, & MacD, 1988) and work relations (Ibarra, 1995; Lincoln & Miller, 1979) to the close bond of marriage (Kalmijn, 1998). It should also be noted that race and ethnicity are among the sociodemographic characteristics that lead to the highest levels of choice homophily (McPherson et al., 2001). For example, a study of interracial friendships in secondary schools found that cross-racial friendships were only one-sixth as likely as choosing a same-race friend (Hallinan & Williams, 1989).

Particularly relevant in the context of resource acquisition online are findings that show that anxiety among members of the racial majority increases when the proportion of racial minorities in the group increases (Stephan & Stephan, 1985). This change in the group composition was shown to cause members of the majority group to experience a threat to their superior status (Abrams & Hogg, 2006; Tajfel, 1974). While the white population has historically been the majority group in the startup funding process, online resource acquisition platforms—such as crowdfunding—democratize access to venture funding (see chapter 2.2) and thus increase the proportion of racial minorities. As a result, this change could lead to a perceived threat to the formerly mighty group of white individuals, explaining why white investors could continue to show a tendency to support white entrepreneurs. However, these findings only illustrate why homophily might occur among majority group members in venture funding, i.e., the white population historically. This does not provide a possible reason for homophily in the dimension race and ethnicity in general. In this light, it is also important to mention a study on racial discrimination in crowdfunding, showing that black entrepreneurs are discriminated against (Younkin & Kuppuswamy, 2018), which also only distinguishes between the white population (majority group) and the black population (minority group). However, a broader analysis is required since there are various races besides only white or black individuals. Therefore, the following section explains why it is conceivable that homophily occurs in general in the dimension race.

Just as for the two sociodemographic characteristics presented in the previous chapters, the theories of self-categorization and social identity (see chapter 2.3.1) could explain the decision-making behavior in online resource acquisition with respect to the dimension race and ethnicity. As explained earlier, individuals classify themselves and others into in-group and out-group members. Furthermore, due to the desire to increase their self-esteem, they tend to react negatively to threats to their social identity, such as discrimination (Avery et al., 2007). This concept seems plausible to also explain why homophily effects might occur with respect to race and ethnicity. According to this concept, investors tend to support projects from entrepreneurs of the same race to prevent discrimination against their race and thus reduce threats to their social identity.

The principle of choice homophily, where individuals choose to interact with similar others solely based on dyadic similarity (Greenberg & Mollick, 2017; Kossinets & Watts, 2009; McPherson & Smith-Lovin, 1987), furthermore provides a general explanation for why investors could support entrepreneurs of the same race in crowdfunding. The idea of choice homophily is further supported by interesting results of several studies showing that people generally recognize faces of their race better than faces of other races (Brigham & Barkowitz, 1978; Malpass & Kravitz, 1969). This same-race bias seems especially relevant in crowdfunding. In this context, it is conceivable that an investor easily recognizes an entrepreneur of the same race based on the profile picture and might therefore, in line with the concept of choice homophily, choose to support the project of an entrepreneur with the same race over the project of an entrepreneur with a different race.

Furthermore, a first study in the crowdfunding context was able to demonstrate racial homophily as investors tend to fund projects from entrepreneurs in the same racial group (Dahlin, Rhue, & Clark, 2019). Thus, according to the theories presented above as well as in line with previous findings,

I propose:

> *Hypothesis 3 (H3): Contributions to entrepreneurs of the same race as the investor occur more frequently than we would expect if investors chose projects for contribution at random (baseline).*

2.3.4. Occupation

Contrary to what has been presented for the previous dimensions, the literature provides arguments opposing the idea of homophily in the dimension occupation, suggesting that investors might notably not support entrepreneurs in the same industry as they are active in. In this context, it is worth mentioning the concept of intragroup competition (Deutsch, 1949; Goldman et al., 1977), which describes how competition arises within a group caused by perceived conflicting goals among the members of the same group. While the concept has previously been used, for example, to explain how members of an organizational department compete for the largest share of a fixed annual budget (Goldman et al., 1977), it also seems appropriate for explaining the behavior of those investors in online resource acquisition who are entrepreneurs themselves. In this context, it seems conceivable that individuals who have themselves launched a campaign on the platform might not want to invest in campaigns within their industry (intragroup), as they might assume that this would reduce the probability of success of their campaign. This seems plausible as investors are generally influenced by the amount of funding that has already been pledged to a project (Kuppuswamy & Bayus, 2018). Thus, the entrepreneur-investor might assume that funding the project of another (potentially competing) entrepreneur in the same industry could make other investors prefer the competitor's campaign, as they could be influenced by the potentially higher sum already pledged to the competitor's project. Ultimately, this could lead entrepreneur-investors not to support projects of other entrepreneurs operating in the same industry so as not to promote competition with respect to their own projects.

However, this is countered by the extensive literature on homophily. According to McPherson et al. (2001), significant homophily has been found in the past for the dimension occupation. For example, in a study in the context of adult friendship choice, similarity in profession was almost equally important and influenced the formation of a tie almost as much as sharing the same gender (Verbrugge, 1977). The sociological finding that individuals choose friends with a similar occupation more often than chance would suggest is explained in part by the principle of status similarity. In this context, a similar social position, for example, in terms of occupation, is valuable because this indicates similar experiences and viewpoints (Verbrugge, 1977). While the previous study examined choices based on similarities in occupation in the friendship context, the logic also seems conceivable in online resource acquisition, as scholars argue that shared experiences and viewpoints also exist in this context. Thereby, especially in crowdfunding, investors often act as campaign creators themselves and can therefore relate to the challenges of their peers, which argues for comradery among this group of people (Colombo, Franzoni, & Rossi–Lamastra, 2015; Groza et al., 2020). In this light, the concept of social capital, generally defined as "the sum of the actual and potential resources embedded within, available through, and derived from the network of relationships possessed by an individual or social unit" (Nahapiet & Ghoshal, 1998, p. 243), is used to explain why investors who are entrepreneurs themselves in general tend to support other entrepreneurs in online venture funding. Social capital includes both the network and the resources mobilized through that network (Bourdieu, 1986) and describes, for example, how shared values and understandings enable individuals to trust each other and thus to cooperate (Nahapiet & Ghoshal, 1998). In this regard, scholars argue for the importance of internal social capital—that is, the contacts established within the crowdfunding platform (Colombo et al., 2015). They argue for a perceived obligation of the entrepreneur to give back to those who helped fund the campaign (specific reciprocity) and a perceived duty to support other campaigns because entrepreneurs are grateful to have received funding in the past (generalized reciprocity). Consistent with this, Colombo et al. (2015) show that investors who are entrepreneurs themselves generally tend to support other entrepreneurs in crowdfunding because they share common challenges, experiences, and a sense of mutual identification.

However, this sense of mutual identification and shared experiences is not limited to the high-level entrepreneur-to-entrepreneur relationship. It is also expected to occur one level deeper in the industry-to-industry relationship, implying that not only do entrepreneurs support each other but that positive homophily effects are also expected between entrepreneurs in the same industry, such as the technology sector. Consistent with the shared values and experiences presented in the social capital literature, this argumentation is further supported by a study in the venture capital context, which shows that industry specialization of the venture capitalists leads to better expertise and thus better investments (Gompers, Kovner, & Lerner, 2009). Since investors do not want to invest in projects that fail, and because of similar experiences and viewpoints, it seems plausible that also in the context of resource acquisition online, investors will invest in campaigns from entrepreneurs that are active in the same industry as themselves.

Consistent with the above argumentation, an initial study in the context of online resource acquisition, specifically crowdlending, was able to show that lenders tend to support projects in the same industry as they are active in (Riggins & Weber, 2017). Thus, following the theories presented above as well as the principle of choice homophily—association due to shared characteristics—I propose:

> *Hypothesis 4 (H4): Contributions to entrepreneurs with the same occupation as the investor occur more frequently than contributions to entrepreneurs with a different occupation than the*

investor.

2.3.5. Location

While all four characteristics presented in chapters 2.3.1 to 2.3.4 are attributed in the literature to the sociodemographic dimensions of homophily that stratify society (McPherson et al., 2001), location is often referred to less as a dimension and more as a source of homophily in the mainstream literature (McPherson et al., 2001). The main reason why geography is considered a source of homophily in the traditional sociological literature is that individuals are simply more likely to interact with individuals who are spatially closer than with individuals who are further away (McPherson et al., 2001). This finding at the same time indicates a high degree of induced homophily with respect to this demographic attribute. In line with this, one study argues, for example, that personal networks are local, suggesting that local contacts are an essential source of routine interactions (Wellman, 1996). This argumentation provides a plausible explanation for traditional sociological research, primarily concerned with forming ties in face-to-face scenarios, with residence inducing a particular frame of possible contacts. In the context of online resource acquisition, however, the situation is fundamentally different, as it is suitable to break the limitation of space. In this case, the geographic location does not seem to be a source of homophily in the traditional sense, as the use of digital platforms removes the direct influence of location on the likelihood of two people interacting. Thus, the location of an individual does not induce the set of possible connections in the online funding process and therefore allows to solely focus on choice homophily with regard to the geographic location. Therefore, it is relevant to investigate whether the location impacts the investors' choice regarding whom to fund. Thus, location is not considered a source but rather a dimension of homophily in my study.

It seems plausible that investors might show homophilous behavior towards entrepreneurs whom they are geographically close to, based on the concept of local bias, which states that investors tend to invest in geographically close companies (Coval & Moskowitz, 1999). Researchers have identified several reasons why investors choose to invest in companies that are close to them. These reasons include, for example, more accessible information about nearby companies, with investors obtaining information from local media or even from local relationships with people from nearby companies (Coval & Moskowitz, 1999, 2001). Other reasons for geographically proximate investments mentioned by scholars are more psychological, such as a desire to fund the local community (Coval & Moskowitz, 1999) or investors simply feeling more comfortable about local firms (Coval & Moskowitz, 1999; Huberman, 2001).

In the past, several studies have shown that the local bias is present in different organizational contexts. For example, studies examining the influence of distance on portfolio choice showed that fund managers in the United States prefer stocks of companies located nearby (Coval & Moskowitz, 1999, 2001). Moreover, consistent with previous findings, an entrepreneurial finance study also showed that venture capital firms exhibit a significant local bias (Cumming & Dai, 2010). Likewise, it seems plausible that geographic proximity is relevant not only in the corporate context but also for private investors, such as in crowdfunding.

In line with this argumentation, the home bias phenomenon, being closely related to local bias, provides further theoretical explanations why investors in online resource acquisition could tend to invest in entrepreneurs whom they are geographically close to. The home bias describes a phenomenon where transactions are conducted among geographically close parties, for example in the same country or state (Lin & Viswanathan, 2016). While home bias has frequently been studied in offline contexts, a study in the context of online auction sites, such as eBay, was able to show that transactions are still more likely to happen between sellers and buyers who are geographically close (Hortaçsu, Martínez-Jerez, & Douglas, 2009). Even though those marketplaces are online, the authors argue that the geographic location is still influential due to location-specific goods like event tickets, shipping charges, or the possibility of direct contract enforcement in case of breach (Hortaçsu et al., 2009).

The results from the studies mentioned above have also been proven in online venture funding, where studies have shown that also in crowdfunding, backers tend to invest locally (Giudici, Guerini, & Rossi-Lamastra, 2018; Guo, Guo, Wang, Wang, & Wu, 2018; Lin & Viswanathan, 2016). Therefore, in line with the above argumentation as well as previous findings, I propose:

> *Hypothesis 5 (H5): Contributions to entrepreneurs with a similar location as the investor occur more frequently than we would expect if investors chose projects for contribution at random (baseline).*

3. Data and Methods

The following chapter provides a description of the data and methods used throughout my analyses. The chapter starts with a brief introduction to the setting of this study (see chapter 3.1), followed by an introduction to the sample, including an explanation of the data collection process (see chapter 3.2). Subsequently, the various variables used in this study are presented (see chapter 3.3). Finally, this chapter ends with the sample's descriptive statistics (see chapter 3.4) and an outline of the statistical methods used in my analyses (see chapter 3.5).

3.1. Setting

To analyze homophily, I selected the crowdfunding platform Indiegogo as the setting of this study. Indiegogo is the second-largest crowdfunding platform after Kickstarter. It belongs to the group of reward-based crowdfunding platforms, where investors receive a reward in return for backing a project with a certain amount of money (Mollick, 2014).

Examples of these rewards—also referred to as *perks* on Indiegogo—include early access to products, the possibility to meet the project creators, or being credited in the supported project (Mollick, 2014).

Founded in 2008 with the mission to empower people to support ideas that matter to them and collectively make those ideas come to life, Indiegogo, with its community of over nine million backers from about 235 countries, has already helped to realize over 800,000 innovative projects (Indiegogo, 2021a). Today, roughly 19,000 campaigns are launched on Indiegogo every month, and about ten million people from all over the world visit the platform during the same time. It is interesting to note that on Indiegogo, women launch 47% of the campaigns exceeding their funding target, which points towards a democratization of the funding process. In line with what has already been presented in chapter 2.2 for crowdfunding platforms in general, Indiegogo as one of those platforms offers the perfect setting to study biases in the choice and decision-making of investors—i.e., choice homophily, the focus of my study. That is because thousands of projects in various categories are offered on Indiegogo as potential investment opportunities for backers from all over the world, where those backers in principle could choose from every possible project, thus lessening induced homophily to a minimum.

In principle, the idea of Indiegogo can be explained as follows. Indiegogo offers a comprehensive platform for entrepreneurs to make their projects available to a large set of potential investors. In return, Indiegogo charges a five percent platform fee on the actual funds raised for the campaign (Indiegogo, 2021d). Entrepreneurs can choose between a fixed funding model (all-or-nothing) or a flexible funding model (keep-what-you-raise). The fixed funding model is used mainly for projects where a strict threshold must be met to realize the project (such as typically in manufacturing). All contributions are returned to the investors if the project does not meet its goal. On the other hand, the flexible funding model, where the entrepreneurs receive all contributions even if the campaign goal was not met, is better suited for projects where no strict minimum financing is required to realize the project. Furthermore, in principle, a campaign on Indiegogo can run for up to 60 days. However, Indiegogo advises a campaign duration of about 40 days, amongst others, for reasons of momentum, engagement, and urgency (Indiegogo, 2021b).

Furthermore, each campaign is assigned to a specific category on Indiegogo. In principle, projects are classified according to the three top-level categories *Tech & Innovation*, *Creative Works*, and *Community Projects* (Indiegogo, 2021c). Each of the three top-level categories comprises a group of more detailed categories. For example, the top-level category *Tech & Innovation* includes categories such as Phones & Accessories, Transportation, Home, or Productivity. The top-level category *Creative Works* contains categories such as Art, Music, or Video Games. Appendix 1 includes a detailed overview of all Indiegogo categories and their allocation to the corresponding top-level categories. Please note that in the following, I will use the term 'top-level categories' when talking about Tech & Innovation, Creative Works, and Community Projects, and I will use the term 'categories' when talking about the more detailed categories.

The individuals who are funding projects on the platform are commonly referred to as *backers* on Indiegogo (Indiegogo, 2021a). When describing the results from my quantitative study, I will refer to this group of people as *backers* or *investors*. The individuals requesting funds are commonly referred to as *entrepreneurs* on Indiegogo (Indiegogo, 2021a). When describing the results from my analyses, I will refer to this group of people as *entrepreneurs* or *founders*.

3.2. Sample

I created a unique portfolio of investments made by backers on the crowdfunding platform Indiegogo. The overall dataset consists of 3,509,077 investments into projects launched between January 2008 and February 2021. Those investments were made by 1,202,233 backers who invested in 171,116 different projects initiated by 147,780 entrepreneurs.

Each observation in my dataset—representing one investment made by one backer—consists of three sections in terms of content. The first section contains various personal information about the backer, such as the name, location, profile picture, personal description, and links to connected social networks. The second section contains various information about the campaign funded, including the category, the amount of funding, the duration of the project, and information about the number of investors, comments, and updates of the project. Finally, the third section contains—similar as for the backer—personal information about the entrepreneur who launched the funded campaign. Therefore, each observation in my dataset represents one dyad, that is, one backer-to-entrepreneur relationship.

I obtained the complete contribution history for all backers in my dataset between January and March 2021. The data collection process can be described as follows. From an initial set of projects launched on Indiegogo, I extracted the unique identifiers (IDs) of all backers who funded those projects using a self-built Python script. This list of roughly 1.2 million different backers represents the actual start of my data collection process. For this purpose, I built a web crawler to collect the portfolios of the individual backers, including all information needed to study the five dimensions of homophily introduced throughout chapter 2.3. To develop my web crawler, I used the open-source framework Scrapy[1]. The process of constructing the portfolio of the backers using my self-built web crawler can be described as follows. First, using the backer IDs as input, I extracted all available personal data about the backers from their Indiegogo profile page (e.g., their name and location, as described above). Second, since Indiegogo profile pages contain not only personal data about the individual but also the history of their

[1] Scrapy is a high-level web crawling framework for Python, used for fast extraction of data from websites (Scrapy, 2021).

activity on the platform, I extracted all contributions made by the backers.

Furthermore, from crawling the funding activity of the backers, I obtained both the IDs of the projects funded and the IDs of the entrepreneurs who launched the corresponding projects. Using this additional information, I collected various data about the projects funded (this data includes, as described above, information about the project category, the funding duration, and the amount of funding). As the last part of my data collection process, I used the same web crawler I already used for the backers for all entrepreneurs who launched the projects to extract all available personal data about the entrepreneurs. By merging the various parts, I finally got to my overall dataset containing about 3.5 million investments made on Indiegogo.

Please note that due to differences in data availability, I created subsamples for my analyses. Those subsamples (i.e., subsets of the overall dataset) were created depending on the availability of the data required for the analysis in the respective dimension. For example, for the analysis in the dimension location, only those observations were considered for which both the investor and the entrepreneur indicated a location on their Indiegogo profile page. The same holds for the other dimensions. For example, only those observations were included in the sample used to analyze racial homophily, where a race was identified for both the backer and the founder. A particular case lies in the creation of the subsample in the dimension occupation. Since the relationship between investors who are entrepreneurs on Indiegogo themselves and the entrepreneurs they fund is examined in this case, the corresponding sample was limited to all those observations where the investor has also launched at least one own campaign. Table 1 shows the number of dyads (# Observations), the number of unique backers (# Backers), the number of unique projects (# Projects), as well as the number of unique founders (# Founders) in the overall dataset as well as for each of the subsamples used in my analyses.

Furthermore, it should be mentioned that while the samples from table 1 contain all investments of all backers (and thus potentially several observations per backer if more than one investment was made by an investor), the various analyses were performed considering one (aggregated) observation per investor. This normalization at the backer level was performed after the calculations required to construct the dependent variables, where the activities of an individual backer were aggregated into special measures per dimension (see chapter 3.3.1). Thus, the sample sizes of my different models are indicated by the number in the column "# Backers" in table 1. However, please note that the final sample sizes were reduced again due to the data availability of the various control variables used in the analyses. Therefore, the final sample sizes are indicated in the figures of the respective models throughout chapter 4.

3.3. Variables

This chapter describes the variables used in my analyses. The chapter starts with a description of the dependent variables (see chapter 3.3.1), followed by the independent variables (see chapter 3.3.2). Afterward, the control variables used in the analyses are presented (see chapter 3.3.3).

3.3.1. Dependent Variables

My basic approach to analyzing homophily in crowdfunding can be summarized as follows. To test for the presence of homophily, I test for deviations of an investor's actual backings from the expected backings—describing what portfolio composition would be expected if the investor chooses ties randomly, i.e., the baseline. This approach is closely related to the example of a person rolling a dice several times. If the die is not loaded (comparable to random choice in my setting), the average of the actual numbers rolled (actual backings) should equal the expected value of the die roll (baseline). However, suppose the average of all the actual numbers rolled significantly deviates from the expected value. In that case, we could suspect a loaded die (i.e., biased choice in my context, which, depending on the interaction with the respective independent variable, indicates homophily).

For further clarification, the approach is now illustrated using a simplified example of the dimension age (compare H1). Let us assume that among all investments made in the crowdfunding universe, the average age of the entrepreneurs funded is 30 years. This average age represents the simplified baseline. Let us now consider the portfolio of one fictitious investor. We expect that the average age of all entrepreneurs funded by this investor is also around 30 years if ties were selected randomly (i.e., if homophily would not be present). Now consider the actual portfolio of this fictitious investor who has backed three projects in the past. One entrepreneur funded by the investor is 19 years old, the second is 20 years old, and the third is 21 years old. Thus, by calculating the mean age of the entrepreneurs in the investor's actual backings, we find that the actual average age of the entrepreneurs funded by the investor is only 20 years. Comparing both numbers indicate a deviation of the backer's actual portfolio from the expected portfolio, whereby the investor, on average, supports younger entrepreneurs than chance would suggest. Depending on whether the backer is younger or older himself, we could ultimately argue for or against homophily.

Therefore, in general, two measures are necessary to construct the dependent variables, which denote the deviation of a backer's actual investments from the baseline. One is the actual backings of an investor, and the other is the expected backings, that is, the baseline. The baseline denotes the expected portfolio of an investor depending on the categories of the projects funded by an investor. The actual backings, on the other hand, represent the actual portfolio of a backer. Thereby, a portfolio contains all investments that an investor has made at the time of the data collection. However, there are slight differences in the type and measurement of the dependent variables, depending on the particular hypothesis of my study considered. The following section thus begins with the description of the dependent variable for the analysis in the dimension age (H1), followed by the dependent variables for the dimensions sex (H2), as well as race (H3).

Table 1: Overview of the overall dataset and different subsamples

Source: Own illustration

Sample	# Observations	# Backers	# Projects	# Founders
Overall	3,509,077	1,202,233	171,116	147,780
Sex (DeepFace)	8,337	3,563	4,315	3,845
Sex (Genderize.io)	2,423,919	945,651	151,364	130,949
Age	8,337	3,563	4,315	3,845
Race	8,337	3,563	4,315	3,845
Location	401,967	109,241	51,771	44,242
Occupation	104,590	27,920	50,250	44,701

Subsequently, the dependent variables for the dimension location (H5) are presented. Please note that the construction of the dependent variable in the dimension occupation (H4) differs from the above-mentioned procedure and is therefore explained at the end of this section.

Age deviance

For the dimension age (H1), the deviation of the actual portfolio of a backer from the expected portfolio is measured using a continuous dependent variable that reflects the extent of deviation (i.e., the delta) of the actual portfolio from the baseline. Thus, the continuous dependent variable (*Age deviance*) measures the extent to which the actual average age of the entrepreneurs in an investors' portfolio deviates from the expected average age of the entrepreneurs (the baseline)[2]. However, it should be noted that the baseline is not calculated simply based on all investments in the entire universe (as in the simplified example mentioned at the beginning of this chapter). Instead, the baseline is refined considering differences in the age of the entrepreneurs depending on the categories of the campaigns funded by an individual investor. This refinement is necessary because the distribution of entrepreneur characteristics differs across the different categories in which a project can be launched on Indiegogo. As shown in Appendix 5, the average age of the entrepreneurs funded in *Energy & Greentech* is almost 40 years. In comparison, the average age of entrepreneurs in the category *Fashion & Wearables* is only 29 years. Thus, to construct a baseline that reflects random choice as realistically as possible, it is necessary to calculate the baseline not only based on the average age of all entrepreneurs in the sample but to refine it depending on the average age of the entrepreneurs in the categories funded by the backer. Using my dataset, which includes the profile pictures of the entrepreneurs scraped from their Indiegogo account, I used the Python framework DeepFace[3] to estimate the age of the entrepreneurs based on their profile pictures.

Sex deviance and Race deviance

In the dimension sex (H2), the deviation of the actual portfolio of a backer from the expected portfolio is measured using a dummy variable (*Sex deviance*) to denote whether the actual share of the female sex in an investor's portfolio is above the expected share. The dependent variable is coded as 1 if the share of backings into female entrepreneurs is above the baseline and 0 otherwise. Something similar but slightly modified also applies to the dimension race (H3). In this dimension, I use a categorical variable (*Race deviance*), which I operationalize via six dummy variables—i.e., one dummy per race—to denote the deviation of a backer's actual portfolio from the baseline. The respective dummy is coded as 1 if an investor's actual share of backings into a particular racial group (such as black entrepreneurs) is above the expected value and 0 otherwise.

In both cases, the actual share of backings per group is measured as the relative frequency of a particular group of entrepreneurs (e.g., female or black entrepreneurs) among all investments of a backer. On the other hand, the baseline describes the expected share of backings per group, considering the differences among the various categories funded by an individual backer. Refining the baseline again makes sense because, as seen in Appendix 5, 32% of all backings in the category *Wellness* are pledged to female entrepreneurs. In comparison, only 3% of all investments in the category *Productivity* are received by female entrepreneurs. The same holds for the dimension race, where, for example, Asian entrepreneurs receive 19% of all investments in the category *Phones & Accessories*. In comparison, Asian entrepreneurs receive only 3% of all backings in the category *Comics*. Just as for the dimension age, I used the Python framework DeepFace to estimate both the sex and the race of the entrepreneurs based on their Indiegogo profile pictures. Additionally, I used the tool genderize.io[4] to determine the sex of the entrepreneurs based on their first names.

[2]The value of the dependent variable *Age deviance* is negative if the actual average age of the entrepreneurs in a portfolio is below the expected average age for the portfolio (indicating that the investor supports younger en-trepreneurs than expected) and vice versa.

[3]DeepFace is a lightweight face recognition and facial attribute analysis framework for Python. The facial attribute analysis allows to determine the age (in integers), sex (female or male) and race (black, white, Asian, Middle Eastern, Indian and Latino Hispanic) of an individual (Python Package Index, 2021a).

[4]Genderize.io is a tool to determine the sex (female or male) of a person given their first name. The tool offers an easy integration for all major languages—such as Python—and compares names with a database of more than 80.000 distinct names to determine the sex of an individual (Genderize.io, 2021).

Actual average distance and Expected average distance (baseline)
Despite using 'deviance' as the dependent variable in the previous dimensions, I use the two measures *Actual average distance and Expected average distance (baseline)* as the dependent variables in the dimension location (H5), to determine whether investors choose ties randomly or tend to invest in geographically close entrepreneurs. Thereby, comparable to what has already been presented above, the actual average distance denotes the average distance between an investor and all the entrepreneurs in the investor's portfolio, and the baseline reflects the expected average distance between an investor and the entrepreneurs, depending on the categories of the projects funded by the investor. Just as for the previous dimensions, refining the baseline considering the differences in the various categories supported by an investor makes sense. That is because Appendix 5 indicates that the average distance between backers and entrepreneurs in the category *Phones & Accessories* is about 6,756 kilometers. In comparison, the average distance between backers and entrepreneurs in the category *Dance & Theater* is only about 1,093 kilometers. To calculate the distances between the backers and entrepreneurs, I first converted the locations of all individuals (which I collected from the Indiegogo profile pages of the backers and founders) into geographic coordinates using the Python framework geopy[5]. Afterward, based on those coordinates, I calculated each dyad's distance (in kilometers) using the Haversine formula[6].

Funded occupation
Finally, it is worth mentioning a difference in the construction of the dependent variable used to analyze homophily in the dimension occupation (H4). While comparing the actual backings to the expected backings in the dependent variables of the previous dimensions, in this dimension it suffices to determine the top-level project category which was funded most frequently by a backer. Hence, I use a categorical dependent variable (*Funded occupation*), which I operationalize via three dummy variables—one dummy per top-level category—to denote the category in which a backer invests most frequently. The respective dummy is coded as 1 if this is the top-level category the investor invests most often into and 0 otherwise. For example, an investor has made five investments in the past, three in projects from the top-level category Tech & Innovation and two in projects from the top-level category Creative Works. The dependent variable would thus denote that the backer invests primarily into campaigns assigned to the top-level category Tech & Innovation, such that the respective dummy is coded as 1 and the other dummies are coded as 0. If no category was uniquely identified as most frequent among the backings of an investor, the backer was removed as an observation from the sample.

[5]Geopy is a Python framework for several popular geocoding services. The framework allows to locate the coor-dinates of addresses, cities, and countries across the globe (Python Package Index, 2021b).

[6]The haversine formula is used to calculate the distance between two points on Earth using their latitude and longitude (Python Package Index, 2021c).

3.3.2. Independent Variables

The various sociodemographic characteristics of the investors make up the key independent variables of my analyses.

Backer age
Since the profile pages on Indiegogo do not include information about an individual's birthday, I also used the facial attribute analysis tool DeepFace to infer the age of the backers from their Indiegogo profile pictures. Thus, to determine whether possible deviations in the investor's actual portfolio from the expected portfolio are due to the investor's preference for similar-age dyads, I use the independent variable (*Backer age*) to denote the age estimate of the investors in whole years.

Backer sex
I also derived data on the sex of the backers by analyzing their profile pictures, using the Python framework DeepFace. Additionally, as for the corresponding dependent variable, I also determined the sex of the investors using the genderize.io tool by algorithmically comparing their first names to the names from the genderize.io database. Thus, to test whether possible deviations in the actual backer portfolio from the baseline are due to a preference for same-sex dyads, I use an independent dummy variable (*Backer sex*) to indicate the sex estimate, where male represents the omitted baseline sex. Thus, the variable is coded as 1 if the investor is female and 0 otherwise.

Backer race
Just as for the previous independent variables, I also derived data on the race of the backers by analyzing their Indiegogo profile pictures using the Python framework DeepFace. I use a categorical independent variable (*Backer race*), which I operationalize via multiple dummy variables to denote an investor's estimated belonging to a particular racial group (i.e., black, white, Asian, Indian, Middle Eastern, or Latino Hispanic). The respective dummy is coded as 1 if the backer belongs to the specific racial group and 0 otherwise, where in general, white represents the omitted baseline race. However, please note that this differs for the various models to test for racial homophily (H3), where the white race, depending on the model, is also part of the analysis.

Backer occupation
To test whether investors show a tendency to fund projects from entrepreneurs with the same occupation as themselves (approximated by the top-level category in my study), I use a categorical independent variable (*Backer occupation*). This variable denotes the top-level category in which the backer is mainly active, i.e., the top-level category in which the backer has launched most own campaigns. For example, suppose a backer has launched three campaigns. Two in Tech & Innovation and one in Creative Works. In that case, the independent variable *Backer occupation* denotes that Tech & Innovation is the category in which the backer mainly launches own campaigns. I used the data about all backers in my dataset who eventually became entrepreneurs on Indiegogo (i.e., those backers who have also launched at least one cam-

paign) to determine the top-level categories in which those entrepreneur-backers started most of their projects. Like for the corresponding dependent variable, those backers were removed from the sample, where no category was identified as most frequent among the campaigns they launched.

3.3.3. Control Variables

Following what prior research in the context of online resource acquisition presented as influential concerning the entrepreneurs in crowdfunding, I include several backer-related variables in my analysis to control for differences in their experience and the quality of the backer profiles. Additionally, my study also controls for the influence of the categories funded by the backers and their sociodemographic characteristics such as age, sex, and race.

First, I control for backer experience with several discrete variables. Following previous studies in the context of crowdfunding (see Groza et al. (2020); Taeuscher, Bouncken, and Pesch (2020); Tan and Reddy (2021)), I control for backer experience by measuring the number of investments made by a backer on Indiegogo (*Number of contributions*). In addition, I control for the number of projects launched by the backer (*Number of own projects*), to control for differences in crowdfunding experience (compare Allison, Davis, Webb, and Short (2017); Cornelius and Gokpinar (2020); Soublière and Gehman (2020); Taeuscher et al. (2020); Tan and Reddy (2021)). Finally, I control for the number of comments a backer has made on the platform (*Number of comments*) to account for differences in platform activity.

Second, I use several control variables to account for differences in the quality of the backer profiles. Because users can link their Facebook account to Indiegogo, it is possible to determine the number of friends of an individual. Thus, I use a discrete variable to denote the number of Facebook friends of a backer (*Number of FBF*), which has previously been used as a potential indicator of the size of an individual's network (exemplary see Anglin, Wolfe, Short, McKenny, and Pidduck (2018); Mollick (2014); Younkin and Kuppuswamy (2018)). Similar to previous studies focusing on the campaign and entrepreneur (see Davis, Hmieleski, Webb, and Coombs (2017); Taeuscher et al. (2020); Younkin and Kuppuswamy (2018)), I control for the quality of the backer profiles using a discrete variable that reflects the number of words a backer uses on the profile page (*Profile word count*). This measure indicates the differences that backers put into creating and maintaining their Indiegogo profile page.

Furthermore, I use a dummy variable to control for the account quality of the backers, indicating whether the Indiegogo account of a backer is verified (*Profile verified*), where the dummy variable is coded as 1 if the account is verified through at least one channel, such as email, and 0 otherwise. Also, comparable to what Scheaf et al. (2018) did at the entrepreneur level, I use dummies to indicate whether a backer has its Facebook account connected to Indiegogo (*Facebook connected*) and whether a backer has its Twitter account linked (*Twitter connected*). For both variables, the dummy is coded as 1 if the backer has the respective account connected and 0 otherwise.

Third, as already explained, projects on Indiegogo are classified according to different categories. Consistent with previous research (see Allison et al. (2017); Cornelius and Gokpinar (2020); Oo, Allison, Sahaym, and Juasrikul (2019); Taeuscher et al. (2020)) and to control for heterogeneity across the various categories, I control for the category in which the backer invests most often. Hence, I include a categorical control variable (*Main category funded*) to denote the category the backer supports most often. Please note that the dependent variable *Funded occupation* (which is used for the analysis in the dimension occupation—see chapter 3.3.1) in principle measures the same. It does so, however, at the top-level category layer. The control variable *Main category funded*, on the other hand, controls at the more granular category layer and is not used in the models to analyze the dimension occupation. Those backers where no category was identified as most frequent among their investments were removed from the sample.

Finally, the respective sociodemographic characteristics used as independent variables (i.e., the age, sex, and race of the backer—see chapter 3.3.2) are vice versa also used as control variables.

3.4. Descriptive Statistics

To provide a comprehensive overview of the dataset, the chapter is structured as follows. First, I present the descriptive statistics based on the overall dataset of roughly 3.5 million backings. This includes, among others, the correlation matrix of the various variables used in my analyses, as well as the baseline table containing the different entrepreneur characteristics per Indiegogo category (which I calculated and used to determine the expected backer portfolios in my analyses). Next follow some descriptive statistics regarding the set of unique backers in my dataset (including, for example, the distribution of sex and race among the backers as well as additional summary statistics). Afterward, the set of unique entrepreneurs in the dataset is briefly described as well. Fourth, I provide a superficial description of the individual campaigns in my dataset, which mainly provides a first impression about the projects funded, including the amounts of money pledged, the project duration, and the like. Finally, and particularly relevant for the analysis in the dimension occupation, this chapter ends with a brief description of specifically those backers who themselves act as entrepreneurs on Indiegogo.

Appendix 2 includes the correlation matrix based on the overall dataset, covering the numerical and dummy variables used in my regression models and provides an overview of the interactions between the various variables. Furthermore, Appendix 3 includes the summary statistics of those variables based on the overall dataset. The statistics show, for example, that among all investments, the average age of the backers is almost 32 years, and the average distance between an investor and the entrepreneurs funded is about 5,045 kilometers. Additionally, table 2 contains an overview of the distri-

bution of sex and race among the backers and entrepreneurs based on all investments.

The table shows that female investors make about 28% of all backings (considering the numbers from genderize.io). On the other side of the dyad, female entrepreneurs receive roughly 22% of all investments (considering the numbers from genderize.io). Furthermore, most investments are made by white backers, making up 68% of all backings. Likewise, white entrepreneurs receive most investments, accounting for 73% of all investments on the receiver side. However, please note that the N for the various tables in this chapter varies depending on whether the specific characteristics were identified for the individuals.

In addition, Appendix 4 provides an overview of the distribution of the roughly 3.5 million investments across the different categories on Indiegogo. Thereby, most backings are pledged to campaigns assigned to the category Film (16% of all investments), followed by Phones & Accessories (12% of all investments), and Health & Fitness (7% of all investments), with two of those most frequently funded categories belonging to the top-level category Tech & Innovation. In the context of different categories, it is worth mentioning the table I calculated and used to determine the different baselines (i.e., the expected portfolios of the individual backers). This table (see Appendix 5) contains the distribution of entrepreneur characteristics, such as their average age or the share per racial group, among all investments in my dataset aggregated per Indiegogo category. Here, its main purpose is to provide an overview of differences among the entrepreneur characteristics in the various categories offered by Indiegogo.

While the previous descriptions were based on all investments, the unique backers in my dataset will now be described. Table 3 provides an overview of some of the main variables describing the backers. The statistics show that the average investor is 31.36 years old and makes just under four contributions on the platform. In addition, an average backer leaves only roughly one comment on the platform and is hardly active as an entrepreneur on Indiegogo, as the average number of own projects of a backer is only 0.03.

Table 4 furthermore provides an overview of the distribution of sex and race among the unique backer population. Interestingly, while according to the numbers from genderize.io, 32% of all backers are female (compare table 4), they account for only 28% of all investments (compare table 2). This pattern suggests that, on average, female investors make fewer investments than men. The same pattern occurs considering the numbers from DeepFace. Furthermore, only 0.7% of the roughly 1.2 million different backers have their Facebook profile connected to their Indiegogo account, and only 0.6% of the investors have their Twitter profile linked. However, about 24% of all backer profiles are verified via at least one channel, such as email.

Although the entrepreneurs are not the focus of my study, some descriptive statistics will nevertheless be presented to provide a complete overview of the dataset. Table 5 includes a summary describing the unique founders. This table indicates that the average entrepreneur is 30.32 years old and makes investments on the platform as well (2.19 contributions on average), but less than the backers (with an average of 3.86 contributions, as presented in table 3). However, with an average of almost two comments, entrepreneurs are more active than backers in this respect. Furthermore, an average entrepreneur launches 1.30 campaigns on Indiegogo.

Table 6 furthermore provides an overview of the distribution of sex and race among the unique entrepreneurs. Comparing those numbers to table 4, it becomes clear that the distribution of the characteristics among the unique entrepreneurs is very similar to the distribution among the unique backers. For example, 69% of all entrepreneurs (compare table 6) and 67% of all backers (compare table 4) are white.

Furthermore, the unique campaigns in my dataset are briefly described in the following. However, please note that this is purely for an initial overview of the dataset. None of these variables are included in my analyses, as the focus of my study is on the backers in crowdfunding. Table 7 summarizes some essential variables describing the campaigns funded. Please be aware that, for the sake of simplicity, table 7 only includes projects launched in the US$ currency, which make up by far the largest share of projects in my dataset.

On average, a campaign is backed by almost 111 investors, has 4.34 updates posted, and contains roughly 24 comments. Interestingly, the average project duration in my dataset reflects the recommendation made by Indiegogo. As already presented in chapter 3.1, Indiegogo recommends a campaign duration of about 40 days. My dataset indicates comparable numbers. Although campaigns on Indiegogo can, in principle, last up to 60 days, the average duration in my dataset is 42.51 days, which is close to the recommendation of Indiegogo.

Furthermore, as described in chapter 3.1, campaigns on Indiegogo can either have a flexible goal or a fixed goal. The vast majority of projects in my dataset have a flexible goal (i.e., 97% of all projects), whereas only 3% of the campaigns in my dataset have a fixed goal. Moreover, 60% of all projects in my sample are assigned to the top-level category Creative Works, followed by Community Projects, which account for 21% of all campaigns. Only 19% of all projects are launched in the top-level category Tech & Innovation. This is interesting because, as presented at the beginning of this chapter, projects from Tech & Innovation are among the most frequently funded projects in my dataset, which indicates that these campaigns receive an above-average number of investments. Appendix 6 contains the fine-grained share of campaigns launched per category. Furthermore, it should also be noted that—as presented in chapter 3.1—Indiegogo states that women launch 47% of campaigns that exceed their funding goal. My sample shows a somewhat similar pattern, where women initiated 37% of the campaigns exceeding their goal.

Finally, those investors who act as entrepreneurs on the platform Indiegogo themselves should briefly be mentioned,

Table 2: Distribution of sex and race among backers and entrepreneurs (all investments)

Source: Own illustration

Characteristic	Share of all investments made	Share of all investments received
Female (DeepFace)	16%	11%
Male (DeepFace)	84%	89%
Female (genderize.io)	28%	22%
Male (genderize.io)	72%	78%
White	68%	73%
Asian	11%	10%
Latino Hispanic	7%	5%
Black	6%	5%
Middle Eastern	6%	6%
Indian	2%	1%

Table 3: Summary statistics of the unique backers

Source: Own illustration

Variable	Min.	Max.	Median	Mean	SD
Backer age	20	67	31	31.36	5.90
Number of contributions	1	3,802	2	3.86	8.79
Number of comments	0	2,003	0	0.85	7.98
Number of own projects	0	235	0	0.03	0.31
Number of FBF	0	5,034	165	371.57	598.69
Profile word count	0	1,548	0	0.49	8.94

Table 4: Distribution of sex and race among the unique backers

Source: Own illustration

Characteristic	Share of Backers
Female (DeepFace)	17%
Male (DeepFace)	83%
Female (genderize.io)	32%
Male (genderize.io)	68%
White	67%
Asian	12%
Latino Hispanic	7%
Middle Eastern	6%
Black	6%
Indian	2%

Table 5: Summary statistics of the unique entrepreneurs

Source: Own illustration

Variable	Min.	Max.	Median	Mean	SD
Founder age	17	70	30	30.32	5.33
Number of contributions	0	1,600	1	2.19	8.36
Number of comments	0	3,016	0	1.97	19.42
Number of own projects	1	228	1	1.30	1.25

Table 6: Distribution of sex and race among the unique entrepreneurs

Source: Own illustration

Characteristic	Share of Founders
Female (DeepFace)	20%
Male (DeepFace)	80%
Female (genderize.io)	38%
Male (genderize.io)	62%
White	69%
Asian	9%
Latino Hispanic	6%
Middle Eastern	6%
Black	8%
Indian	2%

Table 7: Summary statistics of unique campaigns funded in US$

Source: Own illustration

Variable	Min.	Max.	Median	Mean	SD
Number of backers	0	104,448	29	110.71	863.88
Number of updates	0	345	1	4.34	8.61
Number of comments	0	28,210	5	24.18	230.32
Campaign duration (days)	0	2,000	40	42.51	24.13
Total funding (US$)	0	17,595,711	1,911	17,192.46	155,797.84
Campaign start date	2008-01-14	2021-02-05	-	2015-04-17	-
Campaign end date	2010-04-17	2021-04-05	-	2015-06-14	-

which is particularly interesting for my analysis in the dimension occupation. Among the roughly 1.2 million different backers in my dataset, 27,916 have started at least one campaign on Indiegogo. These entrepreneur-backers launch most projects in the top-level category Creative Works (accounting for 69% of all projects launched by them), followed by the top-level category Tech & Innovation (17% of these investors' projects) and Community Projects (14% of all entrepreneur-investor projects). Considering the distribution among the more fine-grained categories, entrepreneur-backers launch most projects in the category Film (35% of all projects), followed by the category Dance & Theater (13% of all projects). These two categories, both from the top-level category Creative Works, therefore account for roughly 48% of all projects launched by these investors. On average, an entrepreneur-backer is 31 years old and is more active than a standard backer, as this person on average launches 1.25 campaigns and contributes to projects 4.81 times. Interestingly, in contrast to what has been presented so far, this investor type shows a more even distribution of sex in my dataset. Just under 42% of all entrepreneur-investors are women, and about 58% are men (considering the numbers from genderize.io). However, white entrepreneur-investors make up the majority here as well, as they account for almost 68% of these investors, followed by Asian entrepreneur-investors, who make up just under 10%.

3.5. Methods

I conducted different quantitative analyses to test the hypotheses I developed throughout chapter 2.3. The differences in the analytical methods used to test my hypotheses arise from the differences in the type and calculation of the dependent variables. I performed regression analyses to study homophily in the dimensions *age*, *sex*, *race*, and *occupation*, and I performed a t-test to analyze homophily in the dimension *location*. In the following, the methods used to test my hypotheses are described in more detail.

In particular, to analyze homophily in the dimension age (H1), I ran an ordinary least squares regression including robust standard errors to tackle heteroscedasticity. The model I ran in this context measures the interaction between the independent variable *Backer age* and the dependent variable *Age deviance* to determine whether investors tend to invest in entrepreneurs they are similar to with respect to age. I used the variance inflation factor to tackle multicollinearity, with the factor being well below 1.7 for all non-categorical variables. Please note that the variance inflation factor was also calculated for all other models in my analyses, whereby the factor is always well below 1.7.

To analyze homophily in the dimensions sex (H2), race (H3), and occupation (H4), I ran logistic regression models. For the analysis in the dimension sex, I ran one logistic regression model. This model measures the interaction be-

tween the independent variable *Backer sex* (i.e., the female sex) and the dependent variable *Sex deviance* (also for the female sex) to investigate whether investors tend to support more entrepreneurs of their sex than suggested by chance. Since the characteristic sex can only be male or female, one model is sufficient to make conclusions for both women and men regarding homophilous choices.

For the analysis in the dimension race, I ran six models (model 3a to 3f), that is, one model per race. In each model, I measure the interactions of the independent variable *Backer race* (i.e., every race apart from the one tested for in the dependent variable) and the respective dependent variable *Race deviance*. The model for the white race, for example, tests whether all races apart from the group of white investors discriminate against white entrepreneurs. If this is true, I could conversely infer that the omitted group of white investors supports more white entrepreneurs than expected, ultimately indicating homophily.

For the analysis in the dimension occupation, I ran three logit models (model 4a to 4c), that is, one model per top-level category. Thereby, I measure the interaction between the categorical independent variable *Backer occupation* and one expression of the dependent variable *Funded occupation* (i.e., Creative Works, Tech & Innovation, or Community Projects). These models are used to analyze whether investors tend to invest mainly in entrepreneurs who are active in the same category as themselves and thus share the same occupation for the sake of this study.

Unlike the previous dimensions, I conducted a one-sided t-test in the dimension location (H5) and compared the *Actual average distance* with the *Expected average distance* to analyze whether backers tend to invest in entrepreneurs whom they are geographically closer to than chance would suggest.

4. Results

The following chapter covers the results of my analyses. Chapter 4.1 presents the results of the models testing hypotheses 1 to 5. Subsequently, chapter 4.2 presents the results of the robustness test. Specifically, this includes the analysis in the dimension sex (H2) using the same model as in chapter 4.1, but with the alternative data from the tool genderize.io to validate the results of the corresponding model in the previous chapter.

4.1. Hypotheses Testing

Dimension Age (H1)

In hypothesis 1, I predicted that contributions to entrepreneurs with a similar age as the investor occur more frequently than we would expect if investors chose projects for contribution at random. That is, investors make homophilous choices as they tend to invest in entrepreneurs whom they are similar to in age. Thus, I predicted a positive relationship between the age of the backer and the age deviance.

Figure 1 presents the result of the ordinary least squares regression model used to test the hypothesis. The results support H1, as the relationship between the age of the backer and the age deviance—indicating how far the actual average age of the entrepreneurs funded by the backer deviates from the expected average age—is positive and significant ($b = 0.10$, $p < .001$). This indicates that for each additional year of backer age, the age deviance also increases by 0.10. The results suggest that, for example, older investors tend to support older entrepreneurs than suggested by chance, thus indicating homophily.

Dimension Sex (H2)

In hypothesis 2, I predicted that contributions to entrepreneurs of the same sex as the investor occur more frequently than we would expect if investors chose projects for contribution at random. Thus, I predicted that backers tend to invest in entrepreneurs with whom they share the same sex. That is, they make homophilous choices with respect to the sociodemographic dimension sex.

Figure 2 presents the result from the logit regression model used to test this hypothesis. The results support H2, as the interaction of the backer-sex with the corresponding sex deviance is positive and significant. Particularly for the model in figure 2, the interaction of the female backer-sex with the female sex deviance—indicating whether a backer supports more female entrepreneurs than expected—is positive and significant ($b = 0.89$, $p < .001$). This result indicates that for female investors, compared to male investors, the odds to fund more female entrepreneurs than expected increase by a factor of 2.43. Since the sociodemographic characteristic sex has only two forms—male and female—it can further be inferred from the results that male investors also tend to support more male entrepreneurs than chance would suggest. Therefore, the results demonstrate that investors, in general, tend to invest in entrepreneurs of the same sex, indicating homophily with respect to this sociodemographic dimension.

Dimension Race (H3)

In hypothesis 3, I predicted that contributions to entrepreneurs of the same race as the investor occur more frequently than we would expect if investors chose projects for contribution at random. Thus, I predicted that investors tend to invest in entrepreneurs with whom they share the same race. That is, they make homophilous choices in the sociodemographic dimension race.

Figures 3 to 8 show the logit regression models used to test for homophily in this dimension. First, the results for the black race are presented (model 3a), followed by the results for the Indian race (model 3b), white race (model 3c), Asian race (model 3d), Latino Hispanic race (model 3e) and finally the results for the Middle Eastern race (model 3f). However, it should already be mentioned that I find mixed results, and therefore, overall, hypothesis 3 is not supported, as significant differences can be observed depending on the specific race considered.

Variables	b	Std Err	p
Independent Variable			
Backer age	0.10***	0.02	0.00
Control Variables			
Backer sex	0.29	0.32	0.37
Backer race		Included	
Main category funded		Included	
Number of own projects	0.01	0.12	0.95
Number of comments	-0.01	0.01	0.39
Number of contributions	0.00	0.01	0.94
Number of FBF	0.00	0.00	0.96
Facebook connected	0.33	0.33	0.32
Twitter connected	-0.33	0.36	0.37
Profile verified	0.29	0.26	0.26
Profile word count	-0.01*	0.00	0.02

*Notes. N = 1,246; *** $p < .001$; ** $p < .01$; * $p < .05$; † $p < .1$*

Figure 1: Ordinary least squares model for the dimension age

Source: Own illustration

Variables	b	Std Err	p
Independent Variable			
Backer sex (Female)	0.89***	0.17	0.00
Control Variables			
Backer age	0.02†	0.01	0.07
Backer race		Included	
Main category funded		Included	
Number of own projects	0.09	0.07	0.22
Number of comments	-0.01	0.01	0.41
Number of contributions	0.02***	0.00	0.00
Number of FBF	0.00	0.00	0.86
Facebook connected	0.08	0.19	0.67
Twitter connected	0.03	0.20	0.87
Profile verified	-0.09	0.15	0.56
Profile word count	0.00	0.00	0.27

*Notes. N = 1,246; *** $p < .001$; ** $p < .01$; * $p < .05$; † $p < .1$*

Figure 2: Logit model for the dimension sex

Source: Own illustration

Model 3a: Black Race

As shown in figure 3, considering the black race deviance—indicating whether a backer supports more black entrepreneurs than expected—the interaction is negative and significant with all backer races except the omitted group of black investors, hence indicating homophilous choices made by black investors. Specifically for white backers, the interaction is negative and significant (b = -1.59, $p < .001$). Thus, for white investors, compared to black investors, the odds to fund more black entrepreneurs than expected change by a factor of 0.20. The same is true for Asian investors, where the interaction is also negative and significant (b = -1.15, $p < .01$). Therefore, for Asian investors, the odds to fund more black entrepreneurs than expected change by a factor of 0.32, which means that being an Asian investor, the probability to fund more black entrepreneurs than expected decreases significantly compared to being a black investor. Also, for the Indian backer group, the interaction is nega-

tive and significant (b = -2.68, p < .05). Thus, for Indian investors, the odds to fund more black entrepreneurs than expected change by a factor of 0.07, compared to black investors. Similar holds for Latino Hispanic investors, where interaction is negative and significant (b = -2.10, p < .001). Thus, for Latino Hispanic investors, the odds to fund more black entrepreneurs than expected change by a factor of 0.12, indicating that the odds to fund more black entrepreneurs than expected significantly decrease if the investor is Latino Hispanic instead of the investor being black. Finally, for the Middle Eastern investors, the interaction is also negative and significant (b = -1.84, p < .001). Therefore, for Middle Eastern investors, the odds to fund more black entrepreneurs than expected changes by a factor of 0.16, compared to the omitted group of black investors.

Since the interaction is negative and significant for all five investor groups in this model, indicating that they all significantly discriminate against black entrepreneurs, it can be conversely concluded that the interaction with the omitted group of black investors is positive and significant. Thus, black investors tend to support more black entrepreneurs than chance would suggest, arguing for homophilous choices. Therefore, considering this model only, the results would support hypothesis 3.

Model 3b: Indian Race

Similar results are found for Indian backers. As shown in figure 4, considering the Indian race deviance—indicating whether an investor supports more Indian entrepreneurs than expected—the interaction with the various racial groups of the investors used in this model is negative and significant. Specifically, for the group of white backers, the interaction is negative and significant (b = -2.56, p < .001). Thus, for white investors compared to Indian investors, the odds of funding more Indian entrepreneurs than expected change by a factor of 0.08, indicating that the probability significantly decreases for white backers. The same is true for black investors, where the interaction is also negative and significant (b = -1.84, p < .05). Thus, for black investors, the odds of funding more Indian entrepreneurs than chance would suggest change by a factor of 0.16, compared to Indian investors. Likewise, for the Asian backer group, the interaction is negative and significant (b = -2.24, p < .01). Therefore, for Asian investors compared to Indian investors, the odds of funding more Indian entrepreneurs than expected change by a factor of 0.11, indicating a decreasing probability. Also, for Latino Hispanic investors, the interaction is negative and significant (b = -2.94, p < .05). Therefore, the odds of funding more Indian entrepreneurs than expected change by a factor of 0.05 if the investor is Latino Hispanic, indicating that the likelihood to fund more Indian entrepreneurs than expected decreases drastically if the investor is Latino Hispanic instead of Indian. Finally, for Middle Eastern backers, the interaction is also negative and significant (b = -2.15, p < .05). Thus, for Middle Eastern investors, the odds of funding more Indian entrepreneurs than expected change by a factor of 0.12, compared to the omitted group of Indian investors.

Since the interaction is negative and significant for all five backer races in this model, it can conversely be inferred that the interaction with the omitted group of Indian investors is positive and significant. That is, Indian investors tend to support significantly more Indian entrepreneurs than expected from random choice. Thus, considering only this model, I find support for my hypothesis on racial homophily.

Model 3c: White Race

Other than what has been presented before, the situation is different for white backers. As shown in figure 5, considering the white race deviance, the interaction with the five backer races included in this model (i.e., every race apart from the group of white investors) is negative, yet not always significant. Thereby, for the group of black backers, the interaction is negative and significant (b = -1.30, p < .001). Thus, for black investors compared to white investors, the odds to fund more white entrepreneurs than expected change by a factor of 0.27, indicating a decreasing likelihood of supporting more white entrepreneurs than expected if the backer is black. The same holds for Asian investors, where the interaction is negative and significant (b = -0.70, p < .001). Thus, for Asian investors, the odds to fund more white entrepreneurs than expected change by a factor of 0.50, meaning that for Asian investors, the odds to fund more white entrepreneurs than expected are only half as high as for white investors. The interaction is also negative and significant for Latino Hispanic backers (b = -0.74, p < .01). Thus, for Latino Hispanic investors, the odds to fund more white entrepreneurs than expected change by a factor of 0.48, indicating a decreasing probability compared to white investors. On the other hand, for Indian investors, the interaction is negative but not significant (b = -0.66, p < .10). Likewise, for Middle Eastern investors, the interaction is also negative but not significant (b = -0.34, p = .19).

Although the interaction is negative for all five investor groups in this model, suggesting that white entrepreneurs tend to make homophilous choices, I still cannot support hypothesis 3 considering only this model. This is because some interactions with the independent variable are not significant in this model. Consequently, I cannot conversely conclude that, in general, white backers support more white entrepreneurs than expected.

Model 3d: Asian Race

As shown in figure 6, considering the Asian race deviance—indicating whether an investor supports more Asian entrepreneurs than expected—the interaction with the different racial groups of investors in this model is negative, yet not always significant. Thereby, for white backers, the interaction is negative and significant (b = -0.78, p < .001). Thus, for white investors compared to Asian investors, the odds of funding more Asian entrepreneurs than expected change by a factor of 0.46, meaning that for a white investor, the likelihood of supporting more Asian entrepreneurs than chance would suggest decreases. The same is true for Middle Eastern investors, where the interaction is negative and

Variables	b	Std Err	p
Independent Variable			
Backer race (White)	-1.59***	0.28	0.00
Backer race (Asian)	-1.15**	0.37	0.00
Backer race (Indian)	-2.68*	1.08	0.01
Backer race (Latino Hispanic)	-2.10***	0.54	0.00
Backer race (Middle Eastern)	-1.84***	0.53	0.00
Control Variables			
Backer sex	0.28	0.24	0.24
Backer age	0.00	0.02	0.79
Main category funded		Included	
Number of own projects	0.07	0.09	0.43
Number of comments	0.00	0.01	0.99
Number of contributions	0.01**	0.01	0.01
Number of FBF	0.00***	0.00	0.00
Facebook connected	-0.33	0.26	0.20
Twitter connected	0.13	0.27	0.62
Profile verified	0.07	0.20	0.71
Profile word count	0.00	0.00	0.90

*Notes. N = 1,246; ***p < .001; **p < .01; *p < .05; †p < .1*

Figure 3: Logit model 3a for the dimension race (Race deviance—Black)

Source: Own illustration

Variables	b	Std Err	p
Independent Variable			
Backer race (White)	-2.56***	0.67	0.00
Backer race (Black)	-1.84*	0.86	0.03
Backer race (Asian)	-2.24**	0.80	0.01
Backer race (Latino Hispanic)	-2.94*	1.19	0.01
Backer race (Middle Eastern)	-2.15*	0.92	0.02
Control Variables			
Backer sex	0.73†	0.43	0.09
Backer age	0.03	0.03	0.35
Main category funded		Included	
Number of own projects	-0.06	0.23	0.80
Number of comments	-0.01	0.02	0.55
Number of contributions	0.02**	0.01	0.01
Number of FBF	0.00*	0.00	0.02
Facebook connected	-0.39	0.52	0.45
Twitter connected	-0.15	0.56	0.79
Profile verified	-0.18	0.40	0.66
Profile word count	0.00	0.01	0.94

*Notes. N = 1,246; ***p < .001; **p < .01; *p < .05; †p < .1*

Figure 4: Logit model 3b for the dimension race (Race deviance—Indian)

Source: Own illustration

Variables	b	Std Err	p
Independent Variable			
Backer race (Black)	-1.30***	0.25	0.00
Backer race (Asian)	-0.70***	0.20	0.00
Backer race (Latino Hispanic)	-0.74**	0.25	0.00
Backer race (Indian)	-0.66†	0.39	0.09
Backer race (Middle Eastern)	-0.34	0.26	0.19
Control Variables			
Backer sex	-0.34*	0.16	0.03
Backer age	0.00	0.01	0.91
Main category funded		Included	
Number of own projects	0.11	0.07	0.11
Number of comments	0.00	0.00	0.26
Number of contributions	-0.01*	0.00	0.02
Number of FBF	0.00**	0.00	0.01
Facebook connected	0.06	0.16	0.71
Twitter connected	0.07	0.17	0.68
Profile verified	-0.11	0.13	0.41
Profile word count	0.00	0.00	0.56

*Notes. N = 1,246; ***p < .001; **p < .01; *p < .05; †p < .1*

Figure 5: Logit model 3c for the dimension race (Race deviance—White)

Source: Own illustration

significant (b = -0.84, p < .05). Thus, for Middle Eastern investors, the odds of funding more Asian entrepreneurs than expected change by a factor of 0.43, indicating that for Middle Eastern investors, the probability of funding more Asian entrepreneurs than expected is less than half of that for the omitted group of Asian investors. On the other hand, the results for the other three backer races included in this model are negative but not significant. For the group of black investors, the interaction is negative but not significant (b = -0.61, p < .10). Likewise, for Indian backers, the interaction is negative, yet not significant (b = -0.28, p = .57). Furthermore, the interaction for the Latino Hispanic investor group is negative but insignificant as well (b = -0.08, p = .82).

Again, the results of this model do not support hypothesis 3. Although the interaction is negative for all five investor races in this model, which would suggest that the interaction of the omitted group of Asian investors with the Asian race deviance is positive, this conclusion is not possible. That is because the interaction with three of five races is not significant. Thus, for the Asian race, it cannot be inferred that, on average Asian investors tend to support more Asian entrepreneurs than chance would suggest. Hence, I do not find support for my hypothesis on racial homophily considering this model only.

Model 3e: Latino Hispanic Race

As shown in figure 7, considering the Latino Hispanic race deviance, the interaction with the five investor races in the model is also negative but not always significant. For white investors, the interaction is negative and significant (b = -0.84, p < .01). Thus, for white backers compared to Latino Hispanic backers, the odds of funding more Latino Hispanic entrepreneurs than expected change by a factor of 0.43. The same is true for black backers, where the interaction is also negative and significant (b = -1.11, p < .05). Thus, for black investors, the odds of funding more Latino Hispanic entrepreneurs than expected change by a factor of 0.33, indicating that for black investors, the likelihood to fund more Latino Hispanic entrepreneurs than expected is only about one-third as high as for Latino Hispanic investors. On the other hand, the results for the three other races included are negative but not significant. For Asian investors, the interaction is negative but not significant (b = -0.69, p < .10). Likewise, for Indian investors, the interaction is also negative but insignificant (b = -0.92, p = .19). Furthermore, the interaction is negative yet not significant for the Middle Eastern investor group (b = -0.79, p < .10).

Therefore, similar to what has been presented for the white and Asian race (models 3c and 3d), the results of this model do not support hypothesis 3. This is because even though the interaction for all five investor races in this model is negative, indicating a positive interaction of the omitted group of Latino Hispanic investors with the Latino Hispanic race deviance, the interaction is not significant for three of the five races in this model. Thus, since not all racial groups included in this model significantly discriminate against Latino Hispanic entrepreneurs, it is not possible

Variables	b	Std Err	p
Independent Variable			
Backer race (White)	-0.78***	0.22	0.00
Backer race (Middle Eastern)	-0.84*	0.39	0.03
Backer race (Black)	-0.61†	0.36	0.10
Backer race (Indian)	-0.28	0.49	0.57
Backer race (Latino Hispanic)	-0.08	0.34	0.82
Control Variables			
Backer sex	0.26	0.20	0.19
Backer age	0.01	0.01	0.49
Main category funded		Included	
Number of own projects	-0.08	0.09	0.40
Number of comments	0.01†	0.00	0.08
Number of contributions	0.01*	0.00	0.01
Number of FBF	0.00	0.00	0.79
Facebook connected	0.05	0.20	0.81
Twitter connected	-0.17	0.22	0.44
Profile verified	0.17	0.16	0.30
Profile word count	0.00	0.00	0.65

*Notes. N = 1,246; ***p < .001; **p < .01; *p < .05; †p < .1*

Figure 6: Logit model 3d for the dimension race (Race deviance—Asian)

Source: Own illustration

Variables	b	Std Err	p
Independent Variable			
Backer race (White)	-0.84**	0.32	0.01
Backer race (Black)	-1.11*	0.52	0.03
Backer race (Asian)	-0.69†	0.40	0.08
Backer race (Indian)	-0.92	0.70	0.19
Backer race (Middle Eastern)	-0.79†	0.48	0.10
Control Variables			
Backer sex	-0.08	0.25	0.74
Backer age	-0.01	0.02	0.58
Main category funded		Included	
Number of own projects	-0.16	0.12	0.18
Number of comments	0.00	0.00	0.59
Number of contributions	0.01†	0.00	0.06
Number of FBF	0.00	0.00	0.34
Facebook connected	0.59*	0.24	0.01
Twitter connected	0.02	0.25	0.95
Profile verified	-0.20	0.20	0.32
Profile word count	0.00	0.00	0.50

*Notes. N = 1,246; ***p < .001; **p < .01; *p < .05; †p < .1*

Figure 7: Logit model 3e for the dimension race (Race deviance—Latino Hispanic)

Source: Own illustration

to conversely conclude that the omitted group of Latino Hispanic investors exhibits a bias to support significantly more Latino Hispanic entrepreneurs than would be expected by chance.

Model 3f: Middle Eastern Race

As shown in figure 8, considering the Middle Eastern race deviance—indicating whether an investor supports more Middle Eastern entrepreneurs than expected—the interaction with the various racial groups of investors included in this model is negative but not significant. Specifically, for white investors, the interaction is negative but not significant ($b = -0.31$, $p = .38$). The same is true for black investors, where the interaction is negative but not significant ($b = -0.33$, $p = .51$). Likewise, for the group of Asian backers, the interaction is negative but insignificant ($b = -0.66$, $p = .16$). The same holds for Indian investors, where the interaction is also negative yet not significant ($b = -0.92$, $p = .27$). Finally, the interaction is negative but insignificant for the group of Latino Hispanic investor as well ($b = -0.88$, $p = .13$).

Since the results for each of the five backer races in this model are not significant, it is not possible to conclude that Middle Eastern entrepreneurs are significantly discriminated against by investors from all but their racial group. Thus, it is also impossible to conversely conclude that Middle Eastern investors make homophilous choices by supporting more Middle Eastern entrepreneurs than expected. Therefore, considering this model only, the results do not support hypothesis 3.

To summarize the analysis of hypothesis 3, I cannot conclude that investors in general tend to support more entrepreneurs of their race than expected. Thus, hypothesis 3 is not supported as I find mixed results with only models 3a (black) and 3b (Indian) supporting the hypothesis and models 3c to 3f not supporting the hypothesis.

Dimension Occupation (H4)

In hypothesis 4, I predicted that contributions to entrepreneurs with the same occupation as the investor occur more frequently than contributions to entrepreneurs with a different occupation than the investor. I predicted that investors who are active as entrepreneurs on Indiegogo tend to invest in entrepreneurs with whom they share the same occupation. That is, they invest in projects of entrepreneurs in the same top-level category as they launch their own projects.

Figures 9 to 11 present the results from the logit regression models used to test hypothesis 4. Each model corresponds to one top-level category on Indiegogo, representing the occupation in the context of this study. First, the results for the top-level category Creative Works are presented (model 4a), followed by the model for the top-level category Tech & Innovation (model 4b). Finally, the results for the top-level category Community Projects (model 4c) are presented. However, it should already be mentioned that the results of all three models support hypothesis 4, suggesting that investors make homophilous choices as they tend to mainly invest in entrepreneurs with whom they share the same occupation.

Model 4a: Funded Occupation—Creative Works

As shown in figure 9, considering Creative Works as the funded occupation in the dependent variable, the interaction with Creative Works as the backer occupation in the independent variable is positive and significant ($b = 3.42$, $p < .001$). This result suggests that for investors who are active in Creative Works, the odds of mainly funding entrepreneurs active in Creative Works increase by a factor of 30.62, compared to investors who are active in Community Projects. Thus, the result of this model supports H4, as investors significantly tend to invest in the same occupation as they are active when considering the top-level category Creative Works.

Model 4b: Funded Occupation—Tech & Innovation

Comparable results are found for the funded occupation Tech & Innovation (see figure 10). Also in this case, the interaction with the corresponding backer occupation is positive and significant ($b = 2.17$, $p < .001$). These results suggest that for investors who are active in the top-level category Tech & Innovation, the odds of mainly funding entrepreneurs who are active in Tech & Innovation increase by a factor of 8.76 compared to investors who are active in the omitted category Community Projects. Thus, the results of this model also support H4, as investors tend to invest in entrepreneurs with the same occupation when considering the top-level category Tech & Innovation.

Modell 4c: Funded Occupation—Community Projects

Also, the third model (see figure 11) supports hypothesis 4. Considering the funded occupation Community Projects as the dependent variable, the interaction with both expressions of the categorical independent variable for the backer occupation included in this model (i.e., Creative Works and Tech & Innovation) is negative and significant. Precisely, the interaction with the backer occupation Creative Works is negative and significant ($b = -3.98$, $p < .001$). Thus, for investors who are active in Creative Works, the odds of mainly funding entrepreneurs in the top-level category Community Projects change by a factor of 0.02, indicating that the probability significantly decreases compared to investors active in the omitted category Community Projects. Similar results are found for the interaction with the backer occupation Tech & Innovation, which is also negative and significant ($b = -2.75$, $p < .001$). This suggests that for investors whose occupation is Tech & Innovation, compared to investors whose occupation is Community Projects, the odds of mainly funding entrepreneurs in the top-level category Community Projects change by a factor of 0.06, indicating a decreasing probability. Since the probability decreases significantly in both cases, it can be conversely concluded that investors active in the omitted category Community Projects exhibit a bias to invest in entrepreneurs active in the same top-level category. Therefore, the results of this model also support hypothesis 4, as investors tend to support those entrepreneurs with whom they share the same occupation with, also in the case of the top-level category Community Projects.

Variables	b	Std Err	p
Independent Variable			
Backer race (White)	-0.31	0.35	0.38
Backer race (Black)	-0.33	0.51	0.51
Backer race (Asian)	-0.66	0.47	0.16
Backer race (Indian)	-0.92	0.83	0.27
Backer race (Latino Hispanic)	-0.88	0.58	0.13
Control Variables			
Backer sex	0.37	0.23	0.11
Backer age	0.02	0.02	0.28
Main category funded		Included	
Number of own projects	-0.08	0.11	0.48
Number of comments	0.00	0.00	0.56
Number of contributions	0.01*	0.00	0.02
Number of FBF	0.00	0.00	0.69
Facebook connected	-0.08	0.25	0.76
Twitter connected	0.34	0.26	0.20
Profile verified	-0.10	0.20	0.60
Profile word count	0.00	0.00	0.76

*Notes. N = 1,246; ***p < .001; **p < .01; *p < .05; †p < .1*

Figure 8: Logit model 3f for the dimension race (Race deviance—Middle Eastern)

Source: Own illustration

Variables	b	Std Err	p
Independent Variable			
Backer occupation (Creative Works)	3.42***	0.28	0.00
Backer occupation (Tech & Innovation)	-0.07	0.32	0.83
Control Variables			
Backer sex	0.20	0.26	0.43
Backer race		Included	
Backer age	0.02	0.02	0.32
Number of own projects	0.08	0.13	0.55
Number of comments	-0.04*	0.02	0.02
Number of contributions	0.01	0.01	0.52
Number of FBF	0.00*	0.00	0.04
Facebook connected	0.13	0.24	0.58
Twitter connected	0.54*	0.27	0.05
Profile verified	-0.40†	0.22	0.06
Profile word count	0.00†	0.00	0.08

*Notes. N = 830; ***p < .001; **p < .01; *p < .05; †p < .1*

Figure 9: Logit model 4a for the dimension occupation (Creative Works)

Source: Own illustration

Dimension Location (H5)

In hypothesis 5, I predicted that contributions to entrepreneurs with a similar location as the investor occur more frequently than we would expect if investors chose projects for contribution at random. Thus, I predicted that the average distance between a backer and the entrepreneurs

Variables	b	Std Err	p
Independent Variable			
Backer occupation (Creative Works)	-1.15***	0.28	0.00
Backer occupation (Tech & Innovation)	2.17***	0.30	0.00
Control Variables			
Backer sex	-0.21	0.27	0.43
Backer race		Included	
Backer age	-0.01	0.02	0.66
Number of own projects	-0.09	0.13	0.50
Number of comments	0.04*	0.01	0.01
Number of contributions	-0.01	0.01	0.68
Number of FBF	0.00**	0.00	0.01
Facebook connected	-0.34	0.25	0.19
Twitter connected	-0.16	0.28	0.56
Profile verified	0.58*	0.23	0.01
Profile word count	0.00	0.00	0.36

*Notes. N = 830; *** $p < .001$; ** $p < .01$; * $p < .05$; † $p < .1$*

Figure 10: Logit model 4b for the dimension occupation (Tech & Innovation)

Source: Own illustration

Variables	b	Std Err	p
Independent Variable			
Backer occupation (Creative Works)	-3.98***	0.36	0.00
Backer occupation (Tech & Innovation)	-2.75***	0.38	0.00
Control Variables			
Backer sex	0.01	0.38	0.98
Backer race		Included	
Backer age	-0.03	0.03	0.30
Number of own projects	0.00	0.21	0.99
Number of comments	-0.01	0.02	0.74
Number of contributions	0.00	0.02	0.82
Number of FBF	0.00	0.00	0.33
Facebook connected	0.29	0.33	0.38
Twitter connected	-0.70†	0.38	0.07
Profile verified	-0.22	0.30	0.46
Profile word count	0.00	0.00	0.23

*Notes. N = 830; *** $p < .001$; ** $p < .01$; * $p < .05$; † $p < .1$*

Figure 11: Logit model 4c for the dimension occupation (Community Projects)

Source: Own illustration

in the portfolio (*Actual average distance*) is significantly less than the average distance for the backer's expected portfolio (*Expected average distance*). That is, investors make homophilous choices with regard to the sociodemographic dimension location.

The result of the one-sided t-test fully supports hypothesis 5, as the test is highly significant ($p < .001$). Thus, I can conclude that the actual average distance between a backer and the entrepreneurs funded is significantly less than the expected average distance of the portfolio. Therefore, my results suggest that investors do not randomly choose entrepreneurs for contribution but instead exhibit a bias to in-

vest in entrepreneurs to whom they are geographically close.

4.2. Robustness Test

As already mentioned, the sex of the individuals (i.e., of both backers and entrepreneurs) was additionally determined during data collection and processing using the genderize.io tool. Thus, I ran the same model as in chapter 4.1, but with the data from genderize.io, to test the robustness of hypothesis 2 and to validate whether the results are reproducible based on the alternative data.

Dimension Sex (H2)

In hypothesis 2, I predicted that contributions to entrepreneurs of the same sex as the investor occur more frequently than we would expect if investors chose projects for contribution at random. Thus, I predicted that backers make homophilous choices with respect to the sociodemographic characteristic sex.

Figure 12 presents the result from the logit regression model, which I ran as the robustness test to measure homophily in the dimension sex. These results also support hypothesis 2, as the interaction of the backer-sex with the corresponding sex deviance is positive and significant. Particularly for the model in figure 12, the interaction of the female backer-sex with the female sex deviance—indicating whether an investor supports more female entrepreneurs than expected—is positive and significant ($b = 0.88$, $p < .001$). This result suggests that for female investors compared to male investors, the odds to fund more female entrepreneurs than expected increase by a factor of 2.40. As already mentioned in chapter 4.1, it can therefore also be concluded that male investors support more male entrepreneurs than chance would suggest. Thus, these results reproduce the findings of the corresponding test of hypothesis 2 in chapter 4.1, implying that both female investors and male investors tend to support entrepreneurs of the same sex, indicating that backers make homophilous choices with regard to the dimension sex.

5. Discussion

In the past, numerous studies across various disciplines have shown the importance of the construct of homophily. However, most studies only considered one dimension of homophily, such as often the sociodemographic characteristic sex. Thus, previous studies neglected the complexity arising from the multidimensionality of this theoretical construct. Taking the first step to do more justice to the complexity by individually analyzing the five sociodemographic dimensions *age*, *sex*, *race*, *occupation*, and *location* based on the same dataset, I make several contributions.

First, I contribute to the broad literature on homophily. While my findings generally underline the importance of the theoretical construct, as various studies have demonstrated in the past, I add to the literature by showing that homophily is not as clear-cut considering more complex characteristics. That is the dimension race in the specific context of my study. Thus, I extend the existing literature by showing that bias depends on the particular sociodemographic characteristic of an individual under consideration. Thereby, for four of the five dimensions of homophily considered in my study, the results reproduce the findings of previous studies, indicating that homophily plays a vital role in individual choices. Therefore, these results will not be interpreted further here. It suffices to mention that I showed that individuals exhibit significant homophilous behavior in the dimensions age, sex, occupation, and location. That is, backers on the crowdfunding platform Indiegogo make homophilous choices in that they tend to invest in entrepreneurs who have a similar age and location as themselves, as well as in those entrepreneurs with whom they share the same sex and occupation.

More interesting are the results of my analyses in the sociodemographic dimension race, which will be interpreted in more detail in the following. In this context, I found mixed results, with significant differences depending on the race considered. In short, my findings suggest homophilous choices from individuals that are black or Indian. However, the same is not valid for white, Asian, Latino Hispanic, or Middle Eastern investors.

Specifically, my findings suggest that black and Indian entrepreneurs are discriminated against by all but their racial group, implying that on the other hand, backers who are black or Indian significantly tend to support more entrepreneurs of their race than chance would suggest, indicating homophilous choices. Furthermore, my findings provide more nuanced results for the white, Asian, Latino Hispanic, and Middle Eastern race. While in general, homophily seems to be relevant here as well, I do not find that all but their racial group significantly discriminate against those groups of entrepreneurs. Hence, I cannot conversely conclude that individuals from those racial groups exhibit significant homophilous behavior. For example, my findings for the white race suggest that Middle Eastern backers do not significantly discriminate against white entrepreneurs (compare model 3c). Similarly, my results show that Latino Hispanic and Indian investors do not significantly discriminate against Asian entrepreneurs (compare model 3d). Furthermore, Indian investors also do not significantly discriminate against Latino Hispanic entrepreneurs (compare model 3e). Finally, my findings suggest that Middle Eastern entrepreneurs are not significantly discriminated against by investors from any racial group (see model 3f). Therefore, I cannot conclude that, in general, white, Asian, Latino Hispanic, and Middle Eastern investors tend to make homophilous choices by funding more entrepreneurs of their race than expected. The above interpretation indicates that my analyses on racial homophily show different results depending on which of the six races is considered. In light of these findings, it could be interpreted as a result of cultural distance, which states that some cultures are more similar than other cultures (Shenkar, 2001), possibly explaining why certain races are discriminated against by some races but not by others.

Second, I contribute to the crowdfunding literature by

Variables	b	Std Err	p
Independent Variable			
Backer sex (Female)	0.88***	0.11	0.00
Control Variables			
Backer age	0.02*	0.01	0.01
Backer race		Included	
Main category funded		Included	
Number of own projects	0.16**	0.06	0.01
Number of comments	0.00	0.00	0.35
Number of contributions	0.01*	0.00	0.03
Number of FBF	0.00	0.00	0.47
Facebook connected	-0.14	0.13	0.30
Twitter connected	0.18	0.14	0.21
Profile verified	-0.15	0.11	0.15
Profile word count	0.01**	0.00	0.00

*Notes. N = 1,905; *** $p < .001$; ** $p < .01$; * $p < .05$; † $p < .1$*

Figure 12: Logit model for the dimension sex—robustness test

Source: Own illustration

putting the investors into the spotlight of research. While previous studies in the crowdfunding context have mainly focused on the characteristics of the campaigns and the entrepreneurs as the most critical elements for success, past studies have neglected the third influential element: the role of the backers in crowdfunding. This gap is furthermore reflected by the scholars' call to also focus on the backers as a success factor in crowdfunding (Bretschneider & Leimeister, 2017; Kuppuswamy & Bayus, 2018). My study is one of the first to bring the investor into the spotlight of crowdfunding research, thus illuminating the component of crowdfunding success that has often been neglected. At the same time, my findings regarding the choice behavior of the backers also offer some valuable insights for crowdfunding entrepreneurs. As an example, in the dimension race, I demonstrate that, on the one hand, from the backers' perspective, only black and Indian backers make homophilous choices. On the other hand, from the entrepreneurial point of view, black and Indian entrepreneurs might be disadvantaged. While my findings do not permit to conclude that black and Indian entrepreneurs, in general, have fundamentally lower chances of crowdfunding success, I can undoubtedly say that investors of all other races discriminate against them. These findings may explain the emergence of crowdfunding platforms specifically for racial minorities, such as the platform FundBLACKFounders[7]. Furthermore, my findings suggest that, on average, younger entrepreneurs might be in a favorable position to achieve their funding goals. Since my results show that backers exhibit homophily in terms of age, and the average backer is about 31 years old (compare chapter 3.4), it could be interpreted that younger entrepreneurs have higher chances of success than entrepreneurs who are at the upper end of the age range. That is because there are simply more young backers in the crowd who tend to back entrepreneurs of similar age.

Third, I empirically contribute by obtaining the data for my analyses from Indiegogo, the second-largest crowdfunding platform. While most of the studies in crowdfunding use data from Kickstarter (exemplary see Greenberg and Mollick (2017); Kuppuswamy and Bayus (2018)), which is the largest crowdfunding platform, my study is one of the few using data from Indiegogo. Thus, my study offers new insights regarding investors' choices with respect to the backer-to-entrepreneur dyad based on a unique dataset gathered from Indiegogo.

The empirical contribution mentioned above is closely related to my fourth and final contribution. By replicating results found on other crowdfunding platforms, I contribute to the literature by providing an additional perspective on whether results found on one particular platform can be generalized across different crowdfunding platforms. That is because scholars have argued that crowdfunding platforms differ, and therefore, "patterns observed on one platform cannot be assumed to generalize to other platforms" (Dushnitsky & Fitza, 2018, p. 1). In this context, with my analyses based on data from Indiegogo, I reproduced previous results found on other platforms for resource acquisition online. For example, I was able to replicate the results of a study in the context of the lending-based microfinance platform Kiva. These findings suggest that lenders tend to fund fundraisers of the same sex and projects within an industry

[7] FundBLACKFounders is a reward based crowdfunding platform for black entrepreneurs founded to overcome the lack of funding allocated to black entrepreneurs (FundBLACKFounders, 2021).

similar to the one the lenders are active in (Riggins & Weber, 2017). Likewise, my results also suggest that investors tend to support entrepreneurs of the same sex and the same occupation. Just to mention a second example, this also holds for a study based on data from Kickstarter. Thereby, scholars were able to show that women support each other in crowdfunding (Greenberg & Mollick, 2017). This finding is also reflected in my study since my models suggest that investors tend to support entrepreneurs of the same sex, thus replicating these results by showing that female investors tend to fund female entrepreneurs.

6. Conclusion

This study sought out to analyze the effects of homophily. While I can, in principle, prove that investors on Indiegogo make homophilous choices, this does not apply to all dimensions of homophily tested for in my study. While investors primarily exhibit homophilous behavior with respect to the sociodemographic characteristics age, sex, occupation, and location, the answer is not as clear-cut in the more complex dimension race. Instead, significant differences are observable, depending on the race considered, where my results suggest that only black and Indian investors make significant homophilous choices. Despite offering valuable insights on how homophily influences the choice of backers in crowdfunding, this work nevertheless has some limitations that offer potential paths for future research. First, I have tried my best to reflect the mechanism underlying the backers' decision-making behavior. However, I cannot look in the investors' minds. Thus, I cannot judge whether the observed behavior was random or based on a conscious decision. Therefore, an exciting possibility for future research is a qualitative study in which detailed interviews are conducted with crowdfunding backers to understand better why they behave the way they do.

Second, there may be a slight bias in my samples, as only those individuals who uploaded a profile picture including their face were considered—which was caused by the fact that the characteristics of the individuals, such as their age and race, were algorithmically inferred from their profile pictures. All those individuals who did not upload a picture including their face were not taken into account in my study. This could cause a slight bias of reality, as scholars have shown that the face of the entrepreneur represents an important factor in the decision-making process of backers, as projects from entrepreneurs with higher facial trustworthiness have significantly more investors and a higher chance of success (Duan, Hsieh, Wang, & Wang, 2020). Future research might therefore, if possible, use alternative means to determine characteristics of individuals, such as the tools genderize.io, nationalize.io[8], and agify.io[9], to potentially have less biased samples, including not only those individuals where a face can be recognized from their profile pictures. Finally, another limitation arises from using the Python framework DeepFace to extract the characteristics of the individuals from their profile pictures. While technological progress nowadays allows a comparatively good estimation of the attributes of individuals based on their faces, these algorithms are not yet perfectly mature, whereby especially the determination of the age based on a face is still challenging.

Besides the topics mentioned above, my study creates additional exciting areas for future research. First, an increasing number of scholars in the field of homophily recently highlighted the need for studies on homophily that consider the multidimensionality of the concept (Block & Grund, 2014; Nahon & Hemsley, 2014). While my research provides a valuable first step in this direction by analyzing the influence of five different sociodemographic characteristics on homophily based on the same dataset gathered from the crowdfunding platform Indiegogo, further research is required to properly do justice to the complexity arising from the multidimensionality of this construct. In this context, a potentially fruitful path for future research is to apply a holistic view on homophily where different dimensions are not only studied individually but rather the interaction of the various dimensions is investigated.

Second, a more detailed analysis regarding the influence of the dimension race on homophily offers another exciting possibility for future research. That is because the dimension race is simply quite complex, with many possible combinations in a dyad. Since my findings suggest significant differences regarding racial homophily depending on the race considered, a detailed study of homophily considering the characteristic race is necessary to properly investigate which racial groups support each other.

Third, a more in-depth analysis regarding homophily in the dimension occupation is an attractive opportunity for future research. While I analyzed this dimension using the three top-level categories available on Indiegogo, future work could replicate the analyses using the more granular set of approximately 30 different categories (see Appendix 1). It would be interesting to determine whether the homophily hypothesis still holds in this case or whether, due to a higher level of complexity resulting from more possible variations, no strict answer can be given, just as I found for the dimension race.

Finally, replicating my results with data from other crowdfunding platforms is an interesting possibility for future research. On the one hand, this would further help to answer the question of platform generalizability (compare chapter 5). On the other hand, finding the same results on other crowdfunding platforms would additionally underline the robustness of the theoretical construct of homophily in the context of online venture funding. In conclusion, it should be said that my work is only a small step to advance research on homophily, especially in the context of crowdfunding, where I put the backers in the spotlight of research.

[8] Nationalize.io is a tool that predicts the nationality of a person given their name (Nationalize.io, 2021)

[9] Agify.io is a tool to predict the age of a person given their name (Agify.io, 2021)

I hope that in the future, scholars will follow this path to further advance research.

References

Abrams, D., & Hogg, M. A. (2006). *Social identifications: A social psychology of intergroup relations and group processes*. Routledge.

Agify.io. (2021). *Overview.* https://agify.io/#overview. (Accessed: 2021-07-18)

Ahuja, G., Polidoro Jr, F., & Mitchell, W. (2009). Structural homophily or social asymmetry? the formation of alliances by poorly embedded firms. *Strategic Management Journal*, *30*(9), 941–958.

Alba, R. D., & Golden, R. M. (1986). Patterns of ethnic marriage in the United States. *Social Forces*, *65*(1), 202–223.

Allison, T. H., Davis, B. C., Webb, J. W., & Short, J. C. (2017). Persuasion in crowdfunding: An elaboration likelihood model of crowdfunding performance. *Journal of Business Venturing*, *32*(6), 707-725.

Anglin, A. H., Wolfe, M. T., Short, J. C., McKenny, A. F., & Pidduck, R. J. (2018). Narcissistic rhetoric and crowdfunding performance: A social role theory perspective. *Journal of Business Venturing*, *33*(6), 780–812.

Aristotle. (1934). Rhetoric. Nichomachean ethics. In *Aristotle in 23 Volumes.* Harvard University Press.

Avery, D. R., McKay, P. F., & Wilson, D. C. (2007). Engaging the aging workforce: The relationship between perceived age similarity, satisfaction with coworkers, and employee engagement. *Journal of Applied Psychology*, *92*(6), 1542–1556.

Becker-Blease, J. R., & Sohl, J. E. (2007). Do women-owned businesses have equal access to angel capital? *Journal of Business Venturing*, *22*(4), 503–521.

Belleflamme, P., Lambert, T., & Schwienbacher, A. (2014). Crowdfunding: Tapping the right crowd. *Journal of Business Venturing*, *29*(5), 1–25.

Belleflamme, P., Omrani, N., & Peitz, M. (2015). The economics of crowdfunding platforms. *Information Economics and Policy*, *33*, 11–28.

Blau, P. M. (1977). *Inequality and heterogeneity: A primitive theory of social structure* (Vol. 7). Free Press.

Block, P., & Grund, T. (2014). Multidimensional homophily in friendship networks. *Network Science*, *2*(2), 189–212.

Bott, H. (1928). Observation of play activities in a nursery school. *Genetic Psychology Monographs*, *4*(1), 44–88.

Bourdieu, P. (1986). *The forms of capital*. Greenwood.

Bretschneider, U., & Leimeister, J. M. (2017). Not just an ego-trip: Exploring backers' motivation for funding in incentive-based crowdfunding. *The Journal of Strategic Information Systems*, *26*(4), 246–260.

Brigham, J. C., & Barkowitz, P. (1978). Do "They all look alike?" The effect of race, sex, experience, and attitudes on the ability to recognize faces. *Journal of Applied Social Psychology*, *8*(4), 306–318.

Byrne, D. E. (1971). *The attraction paradigm* (Vol. 462). Academic Press.

Carland, J. W., Hoy, F., & Carland, J. A. C. (1988). "Who is an entrepreneur?" is a question worth asking. *American Journal of Small Business*, *12*(4), 33–39.

Colombo, M. G., Franzoni, C., & Rossi–Lamastra, C. (2015). Internal social capital and the attraction of early contributions in crowdfunding. *Entrepreneurship Theory and Practice*, *39*(1), 75–100.

Cornelius, P. B., & Gokpinar, B. (2020). The role of customer investor involvement in crowdfunding success. *Management Science*, *66*(1), 452–472.

Cosh, A., Cumming, D., & Hughes, A. (2009). Outside enterpreneurial capital. *The Economic Journal*, *119*(540), 1494–1533.

Coval, J. D., & Moskowitz, T. J. (1999). Home bias at home: Local equity preference in domestic portfolios. *The Journal of Finance*, *54*(6), 2045–2073.

Coval, J. D., & Moskowitz, T. J. (2001). The geography of investment: Informed trading and asset prices. *Journal of Political Economy*, *109*(4), 811–841.

Cumming, D., & Dai, N. (2010). Local bias in venture capital investments. *Journal of Empirical Finance*, *17*(3), 362–380.

Dahlin, L., Rhue, L., & Clark, J. (2019). *Crowdfunding community formation: Fundraiser race and gender homophily*.

Davis, B. C., Hmieleski, K. M., Webb, J. W., & Coombs, J. E. (2017). Funders' positive affective reactions to entrepreneurs' crowdfunding pitches: The influence of perceived product creativity and entrepreneurial passion. *Journal of Business Venturing*, *32*(1), 90–106.

Deutsch, M. (1949). An experimental study of the effects of co-operation and competition upon group process. *Human Relations*, *2*(3), 199–231.

Duan, Y., Hsieh, T. S., Wang, R. R., & Wang, Z. (2020). Entrepreneurs' facial trustworthiness, gender, and crowdfunding success. *Journal of Corporate Finance*, *64*, 1–24.

Dushnitsky, G., & Fitza, M. (2018). Are we missing the platforms for the crowd? comparing investment drivers across multiple crowdfunding platforms. *Journal of Business Venturing Insights*, *10*, 1–10.

Feld, S. L. (1982). Social structural determinants of similarity among associates. *American Sociological Review*, *47*(6), 797–801.

FundBLACKFounders. (2021). *About.* https://www.fundblackfounders.com/about. (Accessed: 2021-07-13)

Genderize.io. (2021). *Overview.* https://genderize.io/#overview. (Accessed: 2021-04-20)

Giudici, G., Guerini, M., & Rossi-Lamastra, C. (2018). Reward-based crowdfunding of entrepreneurial projects: The effect of local altruism and localized social capital on proponents' success. *Small Business Economics*, *50*(2), 307–324.

Goldberg, C. B. (2005). Relational demography and similarity-attraction in interview assessments and subsequent offer decisions: Are we missing something? *Group & Organization Management*, *30*(6), 597–624.

Goldman, M., Stockbauer, J. W., & McAuliffe, T. G. (1977). Intergroup and intragroup competition and cooperation. *Journal of Experimental Social Psychology*, *13*(1), 81–88.

Gompers, P., Kovner, A., & Lerner, J. (2009). Specialization and success: Evidence from venture capital. *Journal of Economics & Management Strategy*, *18*(3), 817–844.

Greenberg, J., & Mollick, E. (2015). Leaning in or leaning on? gender, homophily, and activism in crowdfunding. *Academy of Management Proceedings*, *62*(2), 341–374.

Greenberg, J., & Mollick, E. (2017). Activist choice homophily and the crowdfunding of female founders. *Administrative Science Quarterly*, *62*(2), 341–374.

Greene, P. G., Brush, C. G., Hart, M. M., & Saparito, P. (2001). Patterns of venture capital funding: Is gender a factor? *Venture Capital: An International Journal of Entrepreneurial Finance*, *3*(1), 63-83.

Groza, M. P., Groza, M. D., & Barral, L. M. (2020). Women backing women: The role of crowdfunding in empowering female consumer-investors and entrepreneurs. *Journal of Business Research*, *117*, 432–442.

Guo, L., Guo, D., Wang, W., Wang, H., & Wu, Y. J. (2018). Distance diffusion of home bias for crowdfunding campaigns between categories: Insights from data analytics. *Sustainability*, *10*(4), 1251.

Hallinan, M. T., & Williams, R. A. (1989). Interracial friendship choices in secondary schools. *American Sociological Review*, *54*, 67–78.

Harrison, R. T., & Mason, C. M. (2007). Does gender matter? women business angels and the supply of entrepreneurial finance. *Entrepreneurship Theory and Practice*, *31*(3), 445–472.

Hegde, D., & Tumlinson, J. (2014). Does social proximity enhance business partnerships? Theory and evidence from ethnicity's role in US venture capital. *Management Science*, *60*(9), 2355–2380.

Hortaçsu, A., Martínez-Jerez, F., & Douglas, J. (2009). The geography of trade in online transactions: Evidence from eBay and mercadolibre. *American Economic Journal: Microeconomics*, *1*(1), 53–74.

Huberman, G. (2001). Familiarity breeds investment. *The Review of Financial Studies*, *14*(3), 659–680.

Huston, T. L., & Levinger, G. (1978). Interpersonal attraction and relationships. *Annual Review of Psychology*, *29*(1), 115–156.

Ibarra, H. (1995). Race, opportunity, and diversity of social circles in managerial networks. *Academy of Management Journal*, *38*(3), 673–703.

Indiegogo. (2021a). *About us.* https://www.indiegogo.com/about/our-story. (Accessed: 2021-03-24)

Indiegogo. (2021b). *Choose launch date & deadline.* https://support.indiegogo.com/hc/en-us/articles/205150367. (Accessed: 2021-03-25)

Indiegogo. (2021c). *Explore.* https://www.indiegogo.com/explore/all?project_type=campaign&project_timing=all&sort=trending. (Accessed: 2021-03-25)

Indiegogo. (2021d). *Pricing & fees.* https://learn.indiegogo.com/pricing-and-fees/. (Accessed: 2021-03-25)

Ingram, P., & Morris, M. W. (2007). Do people mix at mixers? structure, homophily, and the "life of the party". *Administrative Science Quarterly*, *52*(4), 558–585.

Jenq, C., Pan, J., & Theseira, W. (2015). Beauty, weight, and skin color in charitable giving. *Journal of Economic Behavior & Organization*, *119*, 234–253.

Kalmijn, M. (1998). Intermarriage and homogamy: Causes, patterns, trends. *Annual Review of Sociology*, *24*(1), 395–421.

Kalmijn, M., & Vermunt, J. K. (2005). Multidimensional homogeneity of social networks: A multilevel analysis of the role of age and marital status. *Social Networks*, *29*, 25–43.

Kandel, D. B. (1978). Homophily, selection, and socialization in adolescent friendships. *American Journal of Sociology*, *84*(2), 427–436.

Kleinbaum, A. M., Stuart, T. E., & Tushman, M. L. (2013). Discretion within constraint: Homophily and structure in a formal organization. *Organization Science*, *24*(5), 1316–1336.

Kortum, S., & Lerner, J. (2000). Assessing the contribution of venture capital. *The RAND Journal of Economics*, *31*(4), 674–692.

Kossinets, G., & Watts, D. J. (2009). Origins of homophily in an evolving social network. *American Journal of Sociology*, *115*(2), 405–450.

Kuppuswamy, V., & Bayus, B. L. (2018). Crowdfunding creative ideas: The dynamics of project backers. In *The economics of crowdfunding.* Cumming, D. and Hornuf.

Lazarsfeld, P. F., & Merton, R. K. (1954). Friendship as a social process: A substantive and methodological analysis. In *Freedom and control in modern society.* Berger, M. and Abel, T. and Page, C. H.

Lin, M., & Viswanathan, S. (2016). Home bias in online investments: An empirical study of an online crowdfunding market. *Management Science*, *62*(5), 1393–1414.

Lincoln, J. R., & Miller, J. (1979). Work and friendship ties in organizations: A comparative analysis of relation networks. *Administrative Science Quarterly*, *24*(2), 181–199.

Loomis, C. P. (1946). Political and occupational cleavages in a Hanoverian village, Germany: A sociometric study. *Sociometry*, *9*(4), 316–333.

Malpass, R. S., & Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of Personality and Social Psychology*, *13*(4), 330–334.

Massolution. (2015). *The crowdfunding industry report, 2015 CF.* https://www.smv.gob.pe/Biblioteca/temp/catalogacion/C8789.pdf.

McPherson, J. M. (1982). Hypernetwork sampling: Duality and differentiation among voluntary organizations. *Social Networks*, *3*(4), 225–249.

McPherson, J. M. (1983). An ecology of affiliation. *American Sociological Review*, *48*(4), 519–532.

McPherson, J. M., & Smith-Lovin, L. (1982). Women and weak ties: Differences by sex in the size of voluntary organizations. *American Journal of Sociology*, *87*(4), 883–904.

McPherson, J. M., & Smith-Lovin, L. (1986). Sex segregation in voluntary associations. *American Sociological Review*, *51*(1), 61–79.

McPherson, J. M., & Smith-Lovin, L. (1987). Homophily in voluntary organizations: Status distance and the composition of face-to-face groups. *American Sociological Review*, *52*(3), 370–379.

McPherson, J. M., Smith-Lovin, L., & Cook, J. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*(1), 415–444.

Mehra, A., Kilduff, M., & Brass, D. J. (1998). At the margins: A distinctiveness approach to the social identity and social networks of underrepresented groups. *Academy of Management Journal*, *41*(4), 441–452.

Mobley, W. H. (1982). Supervisor and employee race and sex effects on performance appraisals: A field study of adverse impact and generalizability. *Academy of Management Journal*, *25*(3), 598–606.

Mollick, E. (2014). The dynamics of crowdfunding: An exploratory study. *Journal of Business Venturing*, *29*(1), 1–16.

Mollick, E., & Robb, A. (2016). Democratizing innovation and capital access: The role of crowdfunding. *California Management Review*, *58*(2), 72–87.

Nahapiet, J., & Ghoshal, S. (1998). Social capital, intellectual capital, and the organizational advantage. *Academy of Management Review*, *23*(2), 242–266.

Nahon, K., & Hemsley, J. (2014). Homophily in the guise of cross-linking: Political blogs and content. *American Behavioral Scientist*, *58*(10), 1294–1313.

Nationalize.io. (2021). *Overview.* https://nationalize.io/#overview. (Accessed: 2021-07-18)

Oo, P. P., Allison, T. H., Sahaym, A., & Juasrikul, S. (2019). User entrepreneurs' multiple identities and crowdfunding performance: Effects through product innovativeness, perceived passion, and need similarity. *Journal of Business Venturing*, *34*(5), 1–49.

Plato. (1968). Laws [Book Section]. In *Plato in twelve volumes* (Vol. 11). Bury translator. Cambridge: Harvard University Press.

Powell, W. W., White, D. R., Koput, K. W., & Owen-Smith, J. (2005). Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences. *American Journal of Sociology*, *110*(4), 1132–1205.

Python Package Index. (2021a). *Project description - DeepFace.* https://pypi.org/project/deepface/. (Accessed: 2021-04-20)

Python Package Index. (2021b). *Project description - Geopy.* https://pypi.org/project/geopy/. (Accessed: 2021-04-20)

Python Package Index. (2021c). *Project description - Haversine.* https://pypi.org/project/haversine/. (Accessed: 2021-04-20)

Richardson, H. M. (1940). Community of values as a factor in friendships of college and adult women. *The Journal of Social Psychology*, *11*(2), 303–312.

Riggins, F., & Weber, D. (2017). Information asymmetries and identification bias in P2P social microlending. *Information Technology for Development*, *23*(1), 107–126.

Scheaf, D. J., Davis, B. C., Webb, J. W., Coombs, J. E., Borns, J., & Holloway, G. (2018). Signals' flexibility and interaction with visual cues: Insights from crowdfunding. *Journal of Business Venturing*, *33*(6), 720–741.

Scrapy. (2021). *Documentation.* https://docs.scrapy.org/en/latest/. (Accessed: 2021-04-23)

Shenkar, O. (2001). Cultural distance revisited: Towards a more rigorous conceptualization and measurement of cultural differences. *Journal of International Business Studies*, *32*(3), 519–535.

Shrum, W., Cheek Jr, N. H., & MacD, S. (1988). Friendship in school: Gender and racial homophily. *Sociology of Education*, *61*(4), 227–239.

Smith, J. A., McPherson, M., & Smith-Lovin, L. (2014). Social distance in the United States: Sex, race, religion, age, and education homophily among confidants, 1985 to 2004. *American Sociological Review*, *79*(3), 432–456.

Sohl, J. E. (2015). *The angel investor market in 2014: A market correction in deal size*.

Soublière, J. F., & Gehman, J. (2020). The legitimacy threshold revisited: How prior successes and failures spill over to other endeavors on Kickstarter. *Academy of Management Journal*, *63*(2), 472–502.

Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In *Advances in experimental social psychology* (pp. 261–302). Berkowitz, L.

Stephan, W. G., & Stephan, C. W. (1985). Intergroup anxiety. *Journal of Social Issues*, *41*(3), 157–175.

Stuart, T. E., & Sorenson, O. (2007). Strategic networks and entrepreneurial ventures. *Strategic Entrepreneurship Journal*, *1*(3-4), 211–227.

Taeuscher, K., Bouncken, R. B., & Pesch, R. (2020). Gaining legitimacy by being different: Optimal distinctiveness in crowdfunding platforms. *Academy of Management Journal*, *64*(1), 149–179.

Tajfel, H. (1974). Social identity and intergroup behaviour. *Social Science Information*, *13*(2), 65–93.

Tan, Y. H., & Reddy, S. K. (2021). Crowdfunding digital platforms: Backer networks and their impact on project outcomes. *Social Networks*, *64*, 158–172.

Thelwall, M. (2009). Homophily in myspace. *Journal of the American Society for Information Science and Technology*, *60*(2), 219–231.

Tsui, A. S., & Gutek, B. A. (1999). *Demographic differences in organizations: Current research and future directions*. Lexington Books.

Tsui, A. S., & O'reilly, C. A. (1989). Beyond simple demographic effects: The importance of relational demography in superior-subordinate dyads. *Academy of Management Journal*, *32*(2), 402–423.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory* [Book]. Basil Blackwell.

Turner, J. C., & Oakes, P. J. (1986). The significance of the social identity

concept for social psychology with reference to individualism, interactionism and social influence. *British Journal of Social Psychology*, *25*(3), 237–252.

Verbrugge, L. M. (1977). The structure of adult friendship choices. *Social Forces*, *56*(2), 576–597.

Verbrugge, L. M., Gruber-Baldini, A. L., & Fozard, J. L. (1996). Age differences and age changes in activities: Baltimore longitudinal study of aging. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *51*(1), S30–S41.

Vismara, S., Benaroio, D., & Carne, F. (2017). Gender in entrepreneurial finance: Matching investors and entrepreneurs in equity crowdfunding. In *Gender and entrepreneurial activity* (pp. 271–288). Link, A. N. and Phillips, V. B.

Wellman, B. (1996). Are personal communities local? a dumptarian reconsideration. *Social Networks*, *18*(4), 347–354.

Wimmer, A., & Lewis, K. (2010). Beyond and below racial homophily: ERG models of a friendship network documented on Facebook. *American Journal of Sociology*, *116*(2), 583–642.

World Bank. (2013). *Crowdfunding's potential for the developing world.* https://www.infodev.org/infodev-files/wb_crowdfundingreport-v12.pdf.

Younkin, P., & Kuppuswamy, V. (2018). The colorblind crowd? founder race and performance in crowdfunding. *Management Science*, *64*(7), 3269–3287.