JUMS

Volume 9, Issue 1, March 2024

JUNIOR MANAGEMENT SCIENCE

Published by Junior Management Science e.V.

ISSN: 2942-1861

Junior Management Science 9(1) (2024) 1100-1122

Junior Management Science

www.jums.academy
ISSN: 2942-1861

How Sustainable Is Private Equity? Unlocking the Impact of Private Equity on Asset-Level Sustainability: An Empirical Investigation

Paul Sunzenauer

Technical University of Munich

Abstract

The debate over the broader impact of the private equity industry has been a contentious topic in the academic literature. While recently, private equity investors have endorsed sustainability in their investment strategies, little is known whether the industry promotes sustainable transformation. This research uses data from the U.S. Environmental Protection Agency on the emission and handling of toxic chemicals in U.S. factories from 1991 to 2021 as a proxy for facility sustainability. The study reveals that, compared to the overall peer group facilities involved in a private equity takeover reduce pollution by 1.55 %-points less and reduce production waste by 1.1 %-points more in the two years after takeover. Further analysis indicates, that with a higher environmental hazard of the underlying chemicals, both the increase in pollution and the decrease in production waste become more pronounced. The study reveals that private equity ownership does not result in enhanced ecological sustainability. Further, the concurrence of the found trends with generally rising costs of both pollution control and raw materials of higher hazards suggest that the private equity business model is only effective in achieving sustainability goals if those are well aligned with financial objectives.

Keywords: impact of private equity; private equity; SRI; sustainability; sustainable finance

1. Introduction

The following chapter first gives an overview of the topic and the motivation for its selection is provided. Then, the significance of the topic to the scientific community is highlighted and a clear objective for the thesis is formulated. Lastly, a brief outline of the thesis is presented, laying out the structure of the upcoming chapters.

1.1. Background and motivation

At the forefront of modern finance, Modern Portfolio Theory[1] has evolved from a purely financial optimization framework to a comprehensive approach that integrates considering environmental, social and governance (ESG) factors into investment decisions, recognizing the importance of investing for both financial and non-financial outcomes. Sustainability goals such as the United Nations Sustainable Development Goals (SDGs) have created significant public and political pressure on the financial industry to steer investment flows towards projects that promote a sustainable future.[2] The ensuant behavior spurred the development of Socially

In the course of this publication, I would like to thank Prof. Dr. Reiner Braun and the Chair of Entrepreneurial Finance of the Technical University Munich for granting me the opportunity to undertake my master's thesis under such an esteemed guidance. Special thanks go to my supervisor, Lukas Hysky, whose insights and expertise greatly enriched my work. His valuable input, creative concepts, and the flexibility he granted me in pursuing my own ideas were essential to the research process. Lastly, I am indebted to all friends and family members who have supported and believed in my vision, ensuring that this journey was not just challenging but also rewarding.

[1] For the evolution of Modern Portfolio Theory see Elton and Gruber (1997, pp. 1750–1758).

[2] See United Nations (2023); A comprehensive synopsis of the Sustainable Development Goals (SDGs), which commonly serve as the basis for many investors' ESG strategies, can be found on the official United Nations website.

DOI: https://doi.org/10.5282/jums/v9i1pp1100-1122

Responsible Investment (SRI), an investment methodology that assimilates social, environmental and governance considerations to the process of making investment decisions.[3] As for the public capital market, economic uncertainty and financial crises are leading individual and institutional investors to express a preference for companies with better ESG ratings.[4] Not least, ESG has also permeated the private markets, and private equity (PE) market participants have begun to incorporate ESG concerns into their investment strategies and are playing an increasingly significant role in the socially responsible investing space.[5] The private equity industry has a unique opportunity to foster transformation of portfolio companies yet an unambiguous perception whether it has capitalized on this potential remains elusive.[6] A frequent challenge for research on this topic is the dearth of available data from private companies immanent due to their private status. Additionally, readily available data on ESG performance is often not useful for academic research. For example, ESG ratings have been identified as highly biased metrics that do not provide a true picture of a company's true impact on its environing economy.[7] The general data scarcity problem results in academic research often limiting the research to special cases, such as public to private deals or reverse buyouts where the data situation is more affluent. However, it has been demonstrated that such narrow analyses faintly generalize to the entire population of PE deals.[8]

Recently, academia has taken an interest in publicly available pollution data from the United States Environmental Protection Agency (EPA) to examine the impact of private equity ownership on environmental pollution. Emanating evidence provides mixed results regarding the impact of private equity ownership on pollution. Most notably, Abraham et al. (2022) find that average pollution is generally reduced after private equity takeover but is already lower before takeover when the PE investor advocates ESG on its website.[9] Findings from Bellon (2020) infer that a positive effect on environmental impact is the case only in the presence of significant liability risks.[10] On the other hand, Shive and Forster (2020) found that private equity is associated with no effect on greenhouse gas releases when controlling for industry, time and location of the portfolio firms, respectively.[11] In light of this uncertainty, this thesis is motivated by the critical value a thorough understanding of the broader economic implications of the private equity business model has. The findings are critical for both the scientific community and policymakers, as well as for targeting in environmental decision-making and the efforts of environmentalists.

To build on the existing findings and to clarify uncertainties within this empirical domain, this study examines the impact of private equity acquisitions on the sustainability of target firms at the asset level. Subsequently, the motivated research question that this thesis aims to address is: *Does PE-takeover lead to an improved level of environmental pollution management in target companies at the facility-level?*

1.2. Relevance of the research within this work

This thesis contributes to the existing body of literature by exploring a relatively new area of environmental impact assessment, that uses raw granular data about industrial processes of facilities and chemical properties of pollutants to link private equity ownership and environmental outcomes. To operationalize the ESG performance of PE-backed assets, a metric using changes in the amount of toxic chemicals as a proxy for asset-level environmental sustainability is drawn from the data of the Toxics Release Inventory (TRI) of the EPA. Prior studies utilizing the TRI have mainly considered pollution quantities without exploiting the granularity of the TRI's environmental impact measurements as it has been done for other areas of economic research.[12] By proceeding with the toxicological data of the chemicals associated with these pollutions the environmental ramifications of a pollution can be derived. The toxicological study of the nature and quantity of pollution emanating from a PE-backed asset thus allows an unequivocal indication of its environmental impact, yet has not been involved in the research of the impact of private equity.

Conclusively, this unveils a clear gap in the current research on the matter of asset-level sustainability within the private equity asset class that this thesis tries to fill from a toxicological point of view.

1.3. Outline of the thesis

The remainder of this thesis is organized as follows: Chapter 2 presents a review of recent literature on ESG and its adoption in the private equity industry. This is followed by a summary of current research that links the business of private equity to the environmental impacts of portfolio companies. The chapter concludes with the development of hypotheses upon which the research in this thesis is presumed. The third section describes the data sources for the research and the methods used. The fourth section examines the results of the empirical analysis, while the fifth section provides a detailed discussion of the results and the limitations of the study. In addition, this last section provides recommendations for further research on the topic.

For this study, it is important to discriminate between the terms ESG and CSR, as they are not used interchangeably. Specifically, ESG pertains to environmental, social and governance issues, which encompass a broad range of social and economic topics. In contrast, CSR specifically refers to a company's actions with respect to ESG issues, which are usually

[3] See Renneboog et al. (2008, p. 1724).
[4] See Bauer et al. (2021, p. 3977).
[5] See Zaccone and Pedrini (2020, p. 5727).
[6] See Crifo and Forget (2013, pp. 22–23).
[7] See Berg et al. (2022, p. 1316).
[8] See Cohn et al. (2014, pp. 469–490).
[9] See Abraham et al. (2022, p. 29).
[10] See Bellon (2020, pp. 28–29).
[11] See Shive and Forster (2020, p. 1315).
[12] See for example Mastromonaco (2015, pp. 54–55) or Bradley C. Karkkainen (2019, p. 116).

(but not necessarily) hardly quantifiable in financial terms.[13] In essence, the degree to which a company considers ESG issues in its strategic decision-making process can determine its level of CSR. Likewise, an investor's approach to integrating environmental, social and governance considerations into his investment framework defines his strategy as socially responsible investment (SRI).[14]

2. Literature review

A growing body of literature comprises the adoption of ESG in finance and private equity. The following sections provide an overview of the topics underlying this research and introduce the context of the subsequent empirical analysis. First, a discussion on the comprehension of ESG in Finance and private equity is given. Then, recent literature on the broader consequence of PE on the economy is reviewed and lastly the hypotheses are elaborated.

2.1. Risk, value and business opportunity: ESG in finance

Environmental, social and governance, short ESG, factors have gained unprecedented importance in today's business landscape and are more salient than ever in investors' decision-making. Within this connotation, the environmental (E) pillar is concerned with mitigating climate change, reducing pollution, and preserving the natural world.[15] The social (S) pillar refers to social equality, human rights protection, and advocating for consumer rights[16] and the governance (G) pillar addresses corporate governance, tax issues, employee rights, and the promotion of fair compensation. Although non-financial in the nature of its objective, ESG may have direct financial implications. For example, regarding the governance component, empirical evidence suggests that employee representation on supervisory boards has a positive impact on firm efficiency and market valuation.[17] The bundle of ESG criteria incorporated as non-financial objectives into strategic considerations determines the aim of the corporate social responsibility, short CSR of a company. Conclusively, actively allocating capital towards firms with *high* CSR qualifies an investment as socially responsible investment, short SRI, which prioritizes social and environmental outcomes alongside financial returns.

In practice, ESG comprises factors that pose potential risks present within an economy (e.g., the risk of rising sea levels due to climate change), where the risks are innately distributed unevenly across market agents in the economy (e.g., a company located on the coast is at much risk due to rising sea levels). Yet the materialization of these risks must be assumed to be transmittable upon economic interaction of the market agents (e.g., the coastal company may be a supplier to companies located inland), to an unknown degree (e.g., the coastal company supplies a special good which might not be easily substituted). In this notion, a firm maximizing its own shareholder value, can decide, whether to consider only business risks that directly jeopardize the company's business model or also to reduce uncertain risks associated with the company's operations. Following the distinction of Knight (1921), ESG at the market agent level thus reflects a confrontation with risks (i.e., the direct ESG risks) and uncertainties (i.e., the uncertain ESG-risks).[18] A CSR strategy, as a risk management strategy that embodies stewardship in addressing both risks and uncertainties, aims to implement operational measures that not only address immediate risk factors for the company, but also mitigate risk factors for other stakeholders. Due to the antiegalitarian distribution of direct ESG risks, a firm executing a such a strategy will generate non-financial utility that benefit the other stakeholders in managing their direct risks. By implementing CSR strategies, organizations can proactively manage their resilience not only to known risks, but also to unknown uncertainties whose occurrence, timing, and impact are unclear but known to exist.[19] As such, ESG considerations can be regarded as an economic solution to internalize the risks of other stakeholders in order to maximize long-term shareholder value. Thus, the implementation of CSR can be viewed as a Coasian solution to problems associated with social costs, that relies on the principle of internalizing externalities.[20] This is critical in developing an effective CSR strategy for companies, as the costs determine the extent to which the strategy can be implemented.

Likewise for an investor, developing an SRI strategy translates to the question of whether to base investment decisions solely on the idiosyncratic risk-return characteristics of potential investments or to tolerate suboptimal financial performance in order to align with his philosophy of impact related to social responsibility.[21] This requires determining his own willingness-to-pay for the implementation of sustainability in the portfolio; if non-financial utility is to be derived from investing in socially responsible companies, then inferior financial performance to non-SRI investments will be accepted. In cases of refractory underperformance, such investors may consider whether adhering to Friedman, who argued that socially responsible investing is less efficient than investing in better performing conventional funds and using some of the returns to support personal beliefs through charitable giving, represents a more efficient way to achieve positive social impact.[22] This trade-off is especially pertinent in the light of numerous examples where SRI funds underperform ethical

[13] See Gillan et al. (2021, p. 101889) for a detailed elaboration of the terminology.
[14] See Renneboog et al. (2008, p. 1724).
[15] See Goodland (1995, pp. 2–8).
[16] See Littig and Griessler (2005, p. 65).
[17] See Fauver and Fuerst (2006, p. 677).
[18] See Jochen Runde (1998, pp. 539–546); For the original discussion on the meaning of risk and uncertainty see Knight (1921).
[19] See Galbreath (2009, pp. 120–122) for an investigation of strategical objectives of CSR strategies.
[20] See Heal (2005, pp. 387–409) for the application of the Coase theorem on CSR in finance.
[21] See Renneboog et al. (2008, p. 1723).
[22] See Friedman (2007, pp. 173–178); The so called Friedman Doctrine is a common objection to the recent trend of CSR. The approach has created

agnostic funds on a risk adjusted basis but at the same time attract greater capital inflows than traditional funds, resulting in increased profits for the fund manager.[23] Here, limited partners of an investment model also face increased agency costs, since the naturally highly heterogeneous objectives of SRI destabilize a fund managers' obligation to pursue high risk adjusted returns.[24] Hence, for an investor embracing SRI, substantial information about the CSR of a company is essential to mitigate information asymmetry and assess the SRI potential. This is especially crucial, since the quality of CSR is not easily observable and investors can be taken for a ride by managers who endorse responsible investment to pander to investor preferences. The relevance of proper information about sustainability aspects of investees has led to the development of various frameworks aimed at promoting sustainable finance. Alongside the United Nations' Sustainable Development Goals, the European Union's Sustainable Finance Disclosure Regulation (SFDR) is a prominent example of such a framework, designed to increase transparency and standardization of ESG reporting requirements for financial market participants.[25] Similarly, initiatives such as the UN Principles for Responsible Investment (PRI) are driving the creation of new guidelines and values for sustainable investing globally.[26]

The increasing market demand for ESG data is frequently supplied with ESG ratings, which aim at providing investment professionals with ESG data and typically focus on relative performance, providing a broad view of the market with comparisons across industries, peers, and companies. Yet Berg et al. (2022) find a questionable correlation between rating methodologies from different ESG-data vendors that was as low as 0.38. According to their research, ratings diverge due to three aspects: scope divergence, which can be seen as a selection bias in the type of data used for the rating; measurement divergence, where indicators are measured differently (i.e., a subjective application of different scientific methodologies); and finally weight divergences, where emphasis is placed on different issues to assess CSR (i.e. the preferential selection of certain ESG factors over others).[27] Overall, this emphasizes the importance of granular, objective data, obtained through reliable methods when assessing a company's interaction with the environment and society.

To summarize, the internalization of social costs associated with externalities related to ESG factors as well as investors' willingness-to-pay for non-financial utility effectively implement sustainability in economy. However, both fail when stakeholders have diverging perceptions of what constitutes ESG. This underscores the imperative of a collective *theory of impact*, in which all stakeholders in an economy are committed to conjointly pursuing non-financial goals.[28]

2.2. The role of ESG in the private equity industry

With the surge of SRI as investment theme in the last decade, limited partners of private equity firms adopted significant non-financial objectives, that the private equity general partner must deliver with its investment case. As a result, general partners are under increasing pressure from limited partners to allocate funds to projects that meet environmental, social and governance (ESG) objectives.[29] An increasing number of private equity firms have thus expanded their corporate missions to engaging all key stakeholders as response to the heightened focus on topics such as climate change, social issues or technology disruption. During the latest renaissance of private equity, proficient management of ESG risks and the pursuit of ESG as a value opportunity have emerged as a singular investment theme utilized by GPs to uncover novel value drivers.[30]

According to Crifo and Forget (2013), SRI renders two main approaches for private equity firms.[31] First, similar to impact investing strategies in the public capital market, private equity investors can use ESG criteria as a risk screening tool, either negatively to exclude companies, particularly in certain industry sectors for ethical or moral reasons, or positively, to actively seek out specific investment opportunities that align with their ESG principles. Indeed, many private equity investors have been using environmental, social, and governance metrics primarily as a risk management tool, with ESG issues integrated mostly as incumbent risk factors – in a 2020 survey, only 40 % of private equity managers consider ESG as a value opportunity.[32] The second approach private equity investors take to SRI according to Crifo and Forget (2013) is an engagement approach. Other than atomistic public financial market participants, PE has the potential to actively promote objectives such as CSR in a portfolio company. By virtue of their controlling stake in the company, PE investors can disrupt managerial entrenchment and thereby mitigate the tendency to maximize short-term value at the expense of long-term value.[33] As elaborated in the previous chapter, long-term financial value demands preparing for ESG-factors. Thus, PE firms will entail ESG considerations because the private equity business model itself fosters CSR as a side product by linking incentives to the long-term profitability (i.e., considering ESG-risks *and* uncertainties) of the firm.[34] In that sense, the shareholder centric rationale of private equity aligns SRI in the pursuit of the maximization of a

an increasing market for charitable giving, where companies offset the environmental damage of their operations by donating a part of their profits, for example with voluntary carbon offset certificates. This is not to be confused with the market for externalities due to Pigouvian taxes such as pollution permits.

[23] See Liang et al. (2022, pp. 1585–1590)
[24] See Renneboog et al. (2008, pp. 1724–1725).
[25] See EUROSIF (2019).
[26] See UN PRI (2019).
[27] See Berg et al. (2022, pp. 1316–1317).
[28] This highlights the importance of well-defined sustainability objectives to work in an economy, such as the science-based targets, COP15 or the Paris Climate Agreement.
[29] See Bian et al. (2022, pp. 3–5).
[30] See Indahl and Jacobsen (2019, pp. 34–36).
[31] See Crifo and Forget (2013, pp. 22–23).
[32] See Zaccone and Pedrini (2020, p. 5730).
[33] See Shleifer and Vishny (1986, pp. 463–472).
[34] See Sørensen and Yasuda (2022, pp. 16–17).

single objective where the internalization of social costs can be assumed to be compensated by the reduction of agency costs manifested in efficiency gains. It is interesting to contend that, along these lines, stricter ESG regulations, which append higher external costs to social and environmental factors, can be expected to act as Pigouvian taxes that further tighten the alignment of SRI goals with the financial objective of the private equity management model.

The adoption of ESG as value creation opportunity requires the general partners to evaluate an additional set of non-financial data to measure their progress toward SRI goals. This necessitates a shift in the governance model private equity firms apply to their portfolio companies, from traditional financial metrics towards soft information about CSR. The existing ambiguity in ESG data as well as the lack of clear standards can be a significant barrier to pursuing SRI strategies.[35] Essentially, the issue centers on the measurement of the externalities produced by portfolio companies and the allocation of those externalities at the fund level.

2.3. The impact of private equity

A private equity firm is typically structured as a partnership in which general partners, on behalf of limited partners, control and actively monitor the board of directors of their portfolio companies. With that, the private equity firm acts as an intermediary between a large, mostly institutional investor base and the private market, thereby occupying a pivotal role in the financing of unlisted, mostly small to medium-sized companies.[36] These companies require significant capital investment to implement ESG considerations operationally, and their inherent risks and information asymmetries make traditional credit funding difficult to obtain.[37] Even in the presence of legal and regulatory frameworks that mandates a firm to internalize a considerable portion of the social costs linked to ESG factors, the expenses associated with environmental protection and social responsibility can elicit insurmountable illiquidity barriers for private companies. Private equity-backed firms are less constrained financially, in principle enabling them to invest more freely in abatement technologies. Consequently, it is argued that such firms exhibit stronger incentives than other privately held firms to reduce pollution levels when facing increased ESG risks. As such, the PE business models is considered as crucial to facilitating the transition to a more sustainable economy.

However, the empirical evidence whether private equity generates non-financial for the broader economy is inconclusive, as prior research has yielded conflicting findings. While at the time of the first private equity boom, Shleifer and Summers (1988) argued that buyouts create shareholder value at the expense of other stakeholders of the firm,[38] this cannot be directly inferred from current research. However, one conjecture from the extant literature is that for the impact on employer welfare pre-deal ownership status plays a crucial role. Workers in private-to-private targets face increased employment due to transferable skillset growth in addition to a better wage growth in the long run,[39] whereas in public-to-private targets, workers performing automatable tasks face a higher risk of unemployment,[40] with older male workers being significantly worse off than their younger counterparts.[41] Also, research has provided evidence of private equity takeovers leading to a reduction of work hazards for employees, thus contributing to improving the governance pillar of ESG by increasing workplace safety.[42] Lastly, for the government, the immanent debt structure of leveraged buyouts in private equity transactions creates negative spillover effects due to interest tax shields,[43] although evidence suggests that targets typically already have high levels of leverage at the time of a buyout, and increases in debt associated with the buyout only tend to be marginally relevant.[44]

In addition, the impact of private equity on the ESG performance of a portfolio company has been shown to be conditional on the industry of the portfolio company and the regulatory regimes under which the business operates. Examples for the positive asset-level effects of private equity takeover can be found in the food industry for example, where private equity buyouts have been shown to improve the quality of customer service and experience. Restaurants are reported to be better maintained after private equity takeover, especially when the private equity GPs have prior industry experience.[45] On the contrary, private equity ownership has been shown to be detrimental to consumers in sectors where intensive government subsidy and economic moats of incumbents can lead to financial incentives being misaligned with the social utility of the business. This is evidenced in the healthcare industry where the impact of private equity ownership results in negative consequences for other stakeholders, particularly in terms of social factors such as affordable, high-quality health care.[46] A recurring observation in the literature is that the impact of private equity on the wider economy and the environment depends on the regulatory and market framework in which private equity portfolio companies operate. Accordingly, in competitive industries where incentives are aligned between stakeholders, private equity buyouts of companies create value for both consumers of the company and its shareholders. Conversely, in more concentrated industries and those reliant on government as a customer base, private equity ownership often leads to the pursuit of profit maximization at the expense of other stakeholders. This di-

[35] See Eccles et al. (2017, pp. 128–132).
[36] See Kaplan and Strömberg (2009, pp. 122–124).
[37] See Kim and Xu (2022, pp. 576–578).
[38] See Shleifer and Summers (1988, pp. 33–68).
[39] See Agrawal and Tambe (2016, pp. 2455–2460).
[40] See Olsson and Tåg (2017, pp. 697–702).
[41] See Antoni et al. (2019, pp. 634–657).
[42] See Cohn et al. (2021, p. 4835).
[43] See Kaplan (1989, pp. 611–623).
[44] See Cohn et al. (2022, p. 284).
[45] See Bernstein and Sheen (2016, p. 2388).
[46] See Atul Gupta et al. (2021, pp. 2–3).

vergence in incentives between investors and consumers can have long-term negative effects on the society, as evidenced by the impact of PE ownership in higher education, where private equity takeovers resulted in a decline of the quality of education while tuition fees increased.[47] As such, it is critical for policymakers to gain a deep understanding of the inherent structures of the private equity model in order to align the outcomes of the business model with desired societal and environmental objectives. Conversely, the presence of misalignments in this investor model tends to be magnified by the incentive power of the private equity management model to maximize financial objectives.[48]

Generally, discerning the impact of PE firms on asset-level operations is difficult, primarily due to the private status of PE-backed companies, which exempts them from mandatory disclosure of financial and non-financial information. In that light, disclosure laws, which are designed to provide the public with information that is not typically included in the normal exchange of goods and services are valuable sources of unbiased information on private companies. These laws, in the US often referred to as "right-to-know" laws, have been deemed necessary in various sectors of the economy. As for the scientific community, right-to-know laws are thus crucial for conducting research. The Toxic Release Inventory from the US Environmental Protection Agency provides such a data source and is one of the most extensive longitudinal data series on facility environmental performance in the United States. EPA's TRI data, with its comprehensive coverage and facility-level information dating back to 1988, offers a valuable tool for examining the connection between economic activities and their environmental impact. A detailed elaboration of the scope and limitations of the TRI database can be found in chapter 3.1 of this work.

2.4. Hypotheses development

As explained above, based on the shareholder theory, ESG considerations are a valuable resource that private equity managers are incentivized to exploit. The conjecture in this thesis is that after a private equity takeover, regardless of the type of deal, the private equity management will seek to identify and address inefficiencies that generate negative externalities in order to curtail the internalized social costs of the asset. In the pollution data captured by the EPA, this should manifest as a discernible decline in the amount of pollution released commencing from the year of the deal.

The observable outcome as lower pollution has been researched by Shive and Forster (2020), in the context of greenhouse gas emissions of US facilities. They find that independent private facilities have lower greenhouse gas emissions than public firms and that this is possibly a result of concentrated ownership. In their research the private equity ownership, in contrast to the private independent ownership, does not result in lower emissions.[49] A more detailed view was taken by Bellon (2020) in the context of the oil and gas industry, who finds that location-specific environmental liability risk is a key driver of differences in the impact of PE ownership on pollution abatement at PE-owned facilities. The absence of such risks results in private equity negatively influencing pollution at the target facility level (inferring that private equity fails to internalize social costs and liability risks and Pigouvian taxes are required to correct the market failure).[50] An industry-agnostic view is taken by Abraham et al. (2022), who use a staggered difference-in-differences design to find that pollution reduction is less likely for portfolio companies of private equity firms with high ESG disclosure than it is for private equity firms with low or no ESG disclosure. In their study, this is due to the fact that such PE firms select already clean firms in the investment process.[51]

To further this research, the question arises as to whether the reduction in pollution following the acquisition by a private equity firm is a) generally an effect of private equity takeover or only present in certain industries and b) significantly increased at PE owned facilities in an all-else-equal scenario. The corresponding null hypothesis is, that the reduced pollution is a reflection of the baseline decline in pollution among TRI facilities over time. Hence, based on the internalization of social costs in a shareholder value theory and in accordance with the previous literature findings, the following statement is hypothesized:

> *Hypothesis 1a: Private equity takeover leads to a decrease in pollution post deal year*

As mentioned above, CSR can include the consideration of both imminent ESG-related risks and uncertainties related to ESG factors. This is particularly relevant for private equity investors who seek to maximize long-term value over short-term gains. To mitigate such risks, private equity management aims to minimize the potential impact of known potentialities for the occurrence of unknown social costs (i.e., "known unknowns"). Given that the amount of hazardous substance handled at a facility is a significant source of environmental pollution, a private equity acquisition should result in a reduction in the amount of hazardous waste generated by the facility. Thus, the following hypothesis is constituted:

> *Hypothesis 1b: Private equity takeover leads to a decrease in production waste post deal year*

Private equity firms can leverage their expertise and overcome information asymmetries to gain a comprehensive understanding of the assets they acquire. With this knowledge,

[47] See Eaton et al. (2020, pp. 4032–4035).
[48] See Sørensen and Yasuda (2022, p. 41).
[49] See Shive and Forster (2020, pp. 1296–1330); Alternatively to the TRI database, EPA's Greenhouse Gas Reporting Program (GHGRP) has captured CO2-emission equivalents since 2010 and was used in their study.
[50] See Bellon (2020, pp. 28–29).
[51] See Abraham et al. (2022, p. 29).

they can identify and mitigate risks and implement measures to address potentially costly issues. As a result, there should be a discernible divergence between pollution associated with highly hazardous substances and those with minimal environmental or social impact. Along these lines, the following hypothesis is constituted:

> *Hypothesis 2a: Private equity ownership results in a greater reduction of highly dangerous pollution compared to less dangerous pollution.*

Similarly, the assessment of materialization risks related to unknown social costs should result in a discernible difference in the amount of production-related waste generated by highly hazardous versus less hazardous substances, resulting in a reduced environmental hazard from the asset. Hence, the following hypothesis is constituted:

> *Hypothesis 2b: Private equity ownership results in a greater reduction of highly dangerous production waste compared to less dangerous production waste.*

3. Methodology

The following section details the methodology used in this thesis. First, the chapter highlights the data sources used to compile the necessary data to construct the sample. Next, the variables of interest and control variables used in the analysis are outlined. Finally, this section elaborates on the empirical models and examines the methods applied in the analysis.

3.1. Data collection and sample reparation

A database of PE transaction data, facility-specific data, and data on environmental pollution is needed to analyse whether the acquisition of a facility's parent company results in a change in its environmental impact. This subsection first explains the deal data source and the resources used to obtain the environmental pollution data. Finally, the procedure used to assemble the final data sample is presented.

3.1.1. Deal data source

The sample of private equity owned firms is drawn from Preqin's Private Equity Database. The Preqin Private Equity Database contains information on PE firms, their funds, portfolio companies linked to the funds and relevant fund performance metrics such as deal date, financial performance indicators, fundraising amounts and exits. Preqin's data is compiled by extracting information from regulatory filings, press releases, the business press and website content.[52] A challenge encountered in utilizing the Preqin database is the prevalence of inaccuracies in the company names of the target companies. This is because Preqin also captures information on real asset deals or acquisitions of businesses units and spin-offs or carve outs. In addition, the target company identifier provided by Preqin is not compatible with other data platforms used for this research. To obtain accurate company identifications, a manual matching process was performed against the Orbis company database of Bureau van Dijk, a comprehensive business data resource on public and private companies.[53] The Orbis BvD IDs were added as company identifier to the Preqin data sample, which was necessary to achieve a consistent match with the parent companies listed by the EPA. The information used from Preqin and Orbis is as of March 2022 and October 2022, respectively.

3.1.2. Facility-level sustainability data source – the TRI program

To measure the asset-level sustainability of private equity transactions, this study utilizes the Toxic Release Inventory from the Environmental Protection Agency of the United States as data source for environmental pollution. The Toxic Release Inventory is a database maintained by the EPA which contains information on the use of certain toxic chemicals by industrial facilities in the United States. Importantly, the TRI also includes relevant facility information, such as location, including exact address and industry sector, as well as the name of the parent company. The database is open to the public under www.epa.gov and can be downloaded or accessed via an API. The jurisdictional basis for the TRI is founded by the Emergency Planning and Community Right-to-Know Act (EPCRA)[54] enacted in 1986 as a response to a severe incident at a chemical facility in West Virginia.[55] Under the Section 313 - EPCRA, all industrial facilities in the US are required to report to the TRI when they meet the following minimum criteria: (i) their operations include the handling, manufacturing, processing or otherwise use of a listed chemical in quantities greater than a threshold during a calendar year (usually 25,000 pounds; 11.34 metric tons of an individual substance), (ii) more than ten full-time workers are employed and (iii) it is classified under a relevant industry sector.[56] At the time of this research, 770 individually listed chemicals and 33 chemical categories were covered by the TRI; the full list is available on the EPA website.[57] The TRI database includes information on the release of these chemicals to the environment (such as through air emissions, water discharges, and land releases), as well as the use, disposal, and treatment of these chemicals. Figure 1 provides an il-

[52] See Preqin (2023).

[53] See Bureau van Dijk Electronic Publishing Ltd (2023).

[54] See "Toxic Chemical Release Reporting: Community Right-To-Know - PART 372" (1988).

[55] See Franklin (1985); The incident occurred at the same type of chemical plant and just eight months after the Bhopal disaster in India, where a cloud of highly toxic methyl isocyanate gas leaked from a Union Carbide chemical plant in Bhopal, India, on Dec. 4th. Thousands of local people died in which is considered as one of the worst industrial disasters in history.

[56] See EPA (2023); For a full list of all covered industry sectors see appendix B.

[57] See EPA (2022a).

lustration of the TRI's tracking of toxic chemicals.[58] For the research in this thesis, information on two reported quantities of toxic chemicals is taken from the TRI database for each facility:

a) The quantities of on-site releases and off-site releases are combined to give the total releases of toxic chemicals associated with a facility's operations which are indicative of a facility's direct environmental impact;
b) The production waste as the amount of toxic chemicals in all non-product outputs generated by the facility which is indicative of the potential of the facility to cause a hazard to the environment. Notably, the total releases are part of the production waste.

Emissions of environmental pollution of a particular facility can be analyzed both as total annual amounts (in pounds) of hazardous chemicals (using the pristine data from the TRI) and as total annual toxic loads using the Risk Screening Environmental Indicators (RSEI) program from the EPA. RSEI processes TRI data to account for the toxicity of a chemical release based on its environmental implications. This is especially relevant when comparing various chemical releases with respect to their environmental hazards.[59] In this thesis, the full RSEI model itself is not used as it is only available for a subset of a few large facilities in the TRI, but the following information from the RSEI program about the total releases and the production waste is obtained:

1) Carcionogen: This boolean variable indicates whether the chemical associated with a reported quantity is considered as carcinogen by the EPA;[60]
2) Persistent Bioaccumulative Toxics (PBT): This boolean variable indicates whether the chemical exhibits a low or no biodegradability and accumulates in living organisms, persistently in adipose tissues of long-living animals i.e., humans.[61]

This research in this thesis innovates and introduces a modified variable based on these indicators to define the *environmental hazard* of a chemical. The *environmental hazard model* (EHM) developed in this thesis is a primitive measure to assess the severity of the pollution when a chemical is released to the environment. The environmental hazard levels are used as pollution-specific control variables which allow a more detailed assessment of the environmental impact of a facility (i.e., a PE-owed asset) on an ordinary scale from 1 to 4. Figure 2 shows the interpretation of the environmental hazard used for the research in this thesis.[62]

The TRI data used for this research included reporting forms processed by the EPA as of October 19, 2022. The data was retrieved on the facility-chemical-year level, meaning that individual quantities for each specific chemical handled at a facility in a reporting year were obtained. As the focus of this study is the impact of private equity takeover on the environmental impact of a facility, the observation of interest is the change in the environmental impact of a facility ocurring in the year of the acquisition. In this context, the relative changes in the amounts of toxic chemicals present in production waste and total releases, respectively are needed. To facilitate this, the total data sample was grouped by year, facility, and hazard level and aggregated by the hazard level. For each facility-hazard-year observation, the difference d_t for a given reporting year t is obtained as shown in Formula 1 by comparing the mean quantities (Y) two years before (*ante*) and two years after (*post*) the reporting year, relative to the mean quantities from the two years preceding (*ante*) the reporting year. This difference is calculated as

$$d_t = \frac{Y_{t+1} + Y_{t+2} - Y_{t-1} - Y_{t-2}}{Y_{t-1} + Y_{t-2}} \quad \text{if } \forall\ Y_{t+z};\ z = \{-2, -1, 1, 2\} \text{ are non-missing} \tag{1}$$

where Y_t denotes the sum of all quantities of toxic chemicals with the same hazard class handled in the facility in the reporting year, normalized to the facility's productivity level for the reporting year under consideration.[63] Normalization allowed controlling for variations in pollution levels attributable to fluctuations in production output, which is necessary to eliminate the effects of increases or decreases in the facility's productivity on the change in quantities handled at the facility. The difference is calculated for the quantities of total releases and for the quantities of production waste individually, henceforth called *Difference Total Releases* and *Difference Production Waste*, respectively. Each difference is calculated only for those reporting years in a facility-hazard group that exist within conjunction of five consecutive reporting years, encompassing two preceding and two subsequent reporting years.

3.1.3. Data preparation and sample construction

The data was prepared in a staged approach, where the deal data sample was first assembled from the Preqin database and then merged with the TRI database to produce the final sample. The initial Preqin deal data sample consisted of 48,232 private equity acquisitions for which Preqin provides transaction information, involving 38,694 unique companies. Of these, 29,652 unique target companies involved in 37,308 transactions were successfully matched manually to their record in the Orbis database. All records with missing information on the year of the transaction were then removed, resulting in 36,082 deals involving 28,945 unique companies. The initial TRI contains data from 18,526 facilities belonging to 4,887 companies. This pristine database was also matched against the Orbis database on

[58] A detailed explanation of the TRI data model can be viewed in appendix C pp. 3–6.
[59] See EPA (2022b).
[60] See EPA (2022b).
[61] See EPA (2022d).
[62] See p. 7 in appendix C for a more detailed explanation of the EHM.
[63] See p. 19 in appendix C for a schematic explanation how the normalized difference was retrieved.

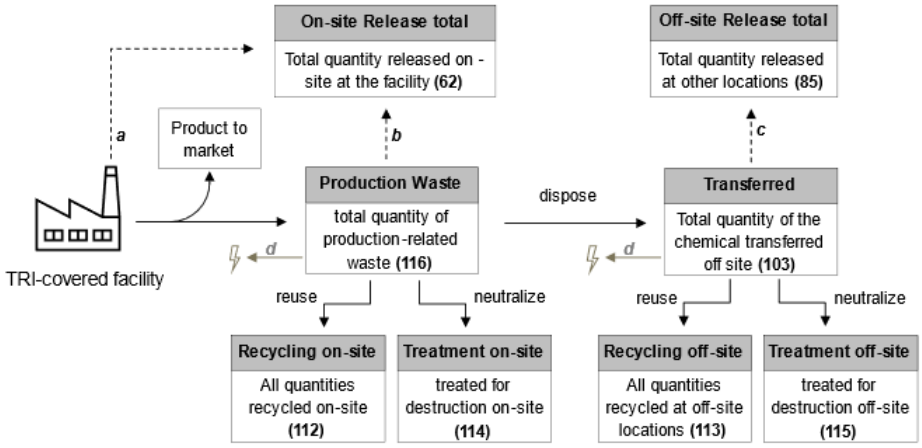


Figure 1: TRI data scope. The figure shows a simplified version of the data notation of the TRI-data points that a facility has to report for any section-313 EPCRA chemical. Numbers in brackets represent data field number in the TRI basic datafile and are used to annotate data for a facility in this thesis too. Dashed errors denote releases to the environment, solid arrows denote transport processes. Sources for pollution are (a) e.g., fugitive or stack air, or (b) e.g., dust or leaching to groundwater while storing, additionally (c) e.g., loss during transportation. While not used in this research, TRI also captures utilization of substance for energy production (d). Also, for (a), (b) and (c): loss of containment as one-time release is covered by the TRI under No. 117, though rarely reported in general. Source: authors own illustration according to EPA (2022c).

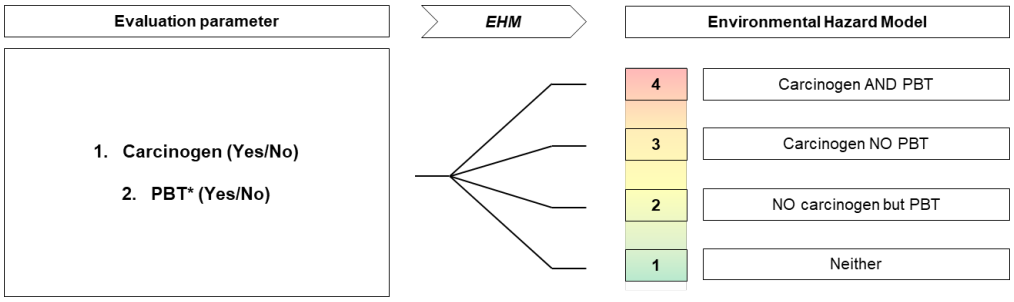


Figure 2: Environmental Hazard Model. The figure shows the interpretation of the environmental hazard of pollution based on RSEI indicators "Carcinogen" and "PBT".

the standardized parent company name[64] using the Orbis search engine, to generate registries with Orbis IDs. To obtain the final sample for the analysis, the TRI sample was then merged to the deal data sample on the Orbis ID which resulted in 330 deals with 219 target companies. A total of 709 TRI-covered facilities were associated with these companies. For all of the 219 TRI-covered target companies, the matched Orbis record was verified manually to prevent false matches of PE-backed companies.

From the final datasample, the following information is obtained: the name of the facility, the verified Orbis ID of the parent company, its location at the state level, its classification in one of 23 industries, the name of the chemical, the amount of production waste and total releases for each chemical used at a facility in a reporting year, respectively, the productivity ratio as the change in productivity from a previous year to the reporting year, and additionally Boolean indicators of whether the chemical is classified as a carcinogen or a persistent bioaccumulative toxic, respectively, and whether the facilities parent company was involved in a transaction during the reporting year. Before assemling the final data sample, all registries corresponding to less than 5 reporting years of a facility as well as registries with missing productivity ratio and or missing quantity for production waste were excluded. The composition of the final sample is summarized in Table 1. The final sample contains data from 1991 to 2021, which resulted in 357,366 facility-year observations in total and thereof 1,054 facility-year observations for facilities associated with an acquired parent company. Since the quantities of toxic chemicals used by a facility in a reporting year are recorded separately for each chemical, the final data sample consisted of 1.45 million facility-chemical-year observations, of which 3,970 observations belonged to facilities involved in a PE transaction. The quantities for production waste and

[64] To enable consolidation of facility-level TRI data with the corresponding parent company, EPA maintains consistent referencing of each company by using the name EPA manually verified. This eliminates typos and variations in names that complicate data aggregation on the parent company level. It also enabled automatic matching by the Orbis search engine with high accuracy.

total releases, respectively, were then aggregated at the facility level by substituting the variables "carcinogen" and "PBT" with the environmental hazard as a model variable. After this, the final sample consisted of 609,916 total and thereof 1,746 deal-related facility-hazard-year observations.

3.2. Variables and empirical models

In this section first, the classification for the independent variable is derived, which is followed by the definition of the dependent variables. Lastly, control variables are established. The chapter then elaborates the construction of the empirical model of this thesis.

3.2.1. Independent variable

A dichotomous independent variable PE is constructed which indicates treatment of an observation (involvement of a facility in a private equity takeover) in a reporting year t. Formula 2 gives the mathematical definition of the independent variable. It is defined as

$$PE = \begin{cases} 1, & \text{the reporting year is a deal year} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

This study assumes that the Preqin database contains all US private equity deals between 1991 and 2021. All observations which exhibit the value PE = 0 constitute the group of untreated observations (= the control group).

3.2.2. Dependent variable

In order to test the hypotheses that private equity takeover leads to a reduction in environmental pollution through reduced total releases and to decreased quantities of toxic chemicals in production waste, the dependent variables are defined as follows:

a) For hypotheses 1a and 2a, the difference d is defined according to Formula 1 for the amount of total releases, distinguishing between hazard levels 1 to 4 for hypothesis 2a.
b) For hypotheses 1b and 2b, the difference d is defined according to Formula 1 for the amount of production waste, distinguishing between hazard levels 1 to 4 for hypothesis 2b.

Importantly, the quantity of toxic chemicals in a reporting year itself is not part of the dependent variable. As shown in Formula 1, the dependent variable considers the quantities of toxic chemical *before* and *after* a reporting year, not *in* a reporting year. This exclusion is critical, as it allows the quantity of toxic chemicals in production waste during year t to be employed as a potential control variable without introducing a logical fallacy into the analysis.

3.2.3. Control variables

Alongside the primary variables of interest, additional control variables are incorporated into the analyses to account for potential confounding factors. Confounding factors are related to facility-fixed effects and pollution-fixed effects. Consequently, the control variables are categorized into the facility-specific control variables and pollution-specific control variables. For the latter, the environmental hazard level h is used. This categorical variable takes into account all effects arising from the potential harm of the chemical when it is released into the environment. This can be exogenous effects, such as tighter regulations for substances exhibiting hazards of highest concern (carcinogenicity, mutagenicity, reproductive toxicity) but also endogenous effects as the cost of precautionary measures increases with higher toxicity of the substances used.

The firm-specific control variables are the facility location on the U.S. state level, the industry sector of the facility, the deal year and the quantity of production waste in the reporting year.

a) The categorical variable facility location k with the 52 U.S. states as categories was included in the model because the sample exhibits significant heterogeneity in geographical distribution and private equity investors exhibit a selection bias for some states.[65] The differences in geographical distribution are highly relevant to the measurement of pollution levels, as state laws regarding environmental protection vary greatly. Previous studies have shown that increased environmental liability risk in a state positively correlates with a better ESG-impact of private equity.[66]
b) The categorical variable industry sector i was included in the model because technological implications due to industry specific processes might potentially influence pollution abatement capabilities in facilities. Additionally, the treatment and control group differed significantly in their distribution of facilities across the industry sectors.[67] For the purposes of this study, facilities were categorized into 23 industries based on their primary NAICS codes.[68]
c) The reporting year t presents an important control variable related to several external effects on the pollution at a facility. External effects include the enactment of environmental protection laws which can greatly distort the treatment effect on pollution change, especially in the location-year combination. Also changes in the TRI reporting requirements from one year to another greatly influence reported quantities and imposes a significant imbalance on comparing assets' inter-year absolute pollution levels.[69]

[65] See p. 12 of appendix C for e graphical representation on the distribution of the facilities and differences thereof between the $PE = 0$ and the $PE = 1$ sample.
[66] See Bellon (2020, p. 2).
[67] See appendix C, pp. 16-17 for all the industrial sectors by their North American Industry Classification System (NAICS) codes present in the sample and a graphical representation of the distribution among industries.
[68] See United States Census Bureau (2022) for the classification of industry sectors on NAICS codes.
[69] See appendix C, pp. 8-11 for the effect of changes in the TRI reporting framework on the reported quantities.

Table 1: Composition of the final data sample. a: The term "in-deal" refers to the count of cases that are associated with parent companies engaged in private equity transactions.

	Total sample	**Thereof in-deal**[a]
Number of facilities	18,526	709
Number of Facility-year observations	357,366	1,054
Number of facility-chemical-year observations	1,445,609	3,970
Number of facility-hazard-year observations	609,916	1,746

d) The size of the facility is an important confounder, yet difficult to exactly assess. The implication of facility size is that large facilities can implement measures more easily, and facilities with high initial pollution can reduce pollution easier. On the other hand, large facilities have a large absolute reduction in pollution, even if the reduction is only a few percentage points. To account for this production waste in the reporting year is used as a proxy for industry size to omit influences of pollution-scale distortions. The measure is imperfect as the ideal measure would be units of production output, but this data is not present. As a compromise, the control variable for the plant size $s = Y(t)$ is introduced as the amount of production waste in year t.

Taken together, the pollution specific and facility-specific control variables constitute a multidimensional vector of characteristics $X_n = \langle f\,(h,k,i,t,s)\rangle$ for each facility-hazard-year observation. In the research design of this thesis, this control vector absorbs salient differences between assets from PE-backed and non-PE-backed facilities.

3.2.4. Empirical models

This thesis follows an adapted notation of Imbens (2004) to develop the model for the estimation of the treatment effect (i.e., the impact of private equity takeover).[70] To begin with, all facility-hazard-year observations are denoted as N cases indexed by n. For each of this case, the differences d_n are observed as dependent variables. Each observed difference for a case n comprises the sum of two components, a population constant baseline difference C_p multiplied by the case-specific vector of characteristics X_n and the effect size E of the treatment effect multiplied by the independent variable PE_n which is 1 if the case has received treatment and 0 otherwise. Formula 3 gives the composition of the observed difference as

$$d_n = E * PE_n + C_p * X_n \tag{3}$$

As the independent variable PE is binary, each case has a pair of potential outcomes for the difference. Accordingly, Formula 4 describes the realized outcome as

$$d_n \equiv d_n(PE_n) = \begin{cases} d_n(1), & if\, PE_n = 1 \\ d_n(0), & if\, PE_n = 0 \end{cases} \tag{4}$$

[70] See Imbens (2004, pp. 5–6).

Substituting $d_n(PE)$ in Formula 4 with Formula 3, the effect size of the treatment can simply be calculated by subtracting the difference $d_n(0)$ from the difference $d_n(1)$. However, as $d_n(0)$ and $d_n(1)$ are never observed for the same case, the effect size must be calculated by subtracting the difference between a treated case n and an untreated case n'. Accordingly, Formula 5 calculates the difference as

$$d_n(PE=1) - d_{n'}(PE=0) = E + C_p * X_n - C_p * X_{n'} \tag{5}$$

Since C_p is not known, for the effect size to be calculated the condition $X_{n'} = X_n$must fulfil. As elaborated in chapter 3.2.3 the vector $X_n$is defined as $X_n = \langle f\,(h,k,i,t,s)\rangle$ and for

$$\begin{aligned} &X_{n'} = X_n \Longleftrightarrow X_{h,k,i,t,s} = X_{h',k',i',t',s'} \\ &\text{and thus for } \{h = h', k = k', i = i', t = t',\ s\ = s'\} \\ &\Longrightarrow n \overset{\text{def}}{=} n';\ \forall n, n' \in N \end{aligned} \tag{6}$$

Under the condition of Formula 6 formula 5 can be written as shown in Formula 7 with

$$\begin{aligned} &d_{h,k,i,t,s}(PE=1) - d_{h,k,i,t,s}(PE=0) \\ &= E + C_p * (X_{h,k,i,t,s} - X_{h,k,i,t,s}) \equiv E \end{aligned} \tag{7}$$

Conclusively, as shown in Formula 6 the effect size of the treatment (the impact of private equity takeover) on the change in quantities of toxic chemicals can be estimated by the difference between two cases (facilities) which are similar in terms of their control variables. The estimation is done for the amount of toxic chemicals production waste and the amount of toxic chemical as total releases separately.

3.3. Empirical methods

To estimate the effect size of the treatment effect on the treated under the aforementioned empirical model, it is necessary to identify an untreated control observation for each treated observation that is similar in terms of confounding factors (i.e., control variables). Subsequently, the difference between the treated and control observations can be calculated. Otherwise, the difference in outcomes could simply be the result of the continuation of pre-existing different trends in the facilities, where the trends are caused or at least related to the confounding factors. To facilitate this, two different methods have been used for matching to fulfil the condition in Formula 6, i.e., to find pairs of cases which are balanced in terms of their vectors of characteristics. In the following, first the propensity score matching method is explained. Since

the matching with the propensity scores was of poor quality, a second matching method was developed which is specifically tailored to the data types present in the sample used for this research.

3.3.1. Probabilistic matching: Propensity Score Matching

In the first approach to investigate whether firms undergoing private equity buyouts achieve lower pollution levels *after* takeover, the Propensity Score Matching (PSM) was employed to create a group of control firms. The propensity score of a subject is the calculated probability of this subject for receiving the treatment conditional on a set of characteristics other than the one being tested for (i.e., the control variables).[71] The selected control variables on which a propensity score for each subject is calculated on must effectively characterize idiosyncratic properties of the subject to avoid overfitting of the propensity score matching.[72] The selection process should be grounded in sound reasoning and take into consideration the extent to which treated and untreated groups differ with respect to each control variable. For the latter, the identification of confounders for the matching was based on the difference between the treated and the untreated group in the context of the respective confounder. Austin (2011b) introduces a measure for the difference for continuous variables as shown in Formula 7.[73] This standardized difference is calculated as

$$d_{st} = \frac{\overline{x}_{treated} - \overline{x}_{untreated}}{\sqrt{\frac{s^2_{treated} + s^2_{untreated}}{2}}} \tag{8}$$

where $\overline{x}$ denotes the mean of treated and untreated samples, respectively, and $s^2_{treated}$ and $s^2_{untreated}$ are the corresponding variances of the treated and untreated sample, respectively. For the categorical variables industry, location and hazard level, the absolute difference in proportions was estimated based on an adapted approach from Austin (2010) as shown in Formula 6.[74] Consequently, in this thesis the difference is calculated as

$$d_{abs} = \frac{1}{2}\sum_{i=1}^{C} \left| p_{treated,c} - p_{control,c} \right| c \in C \tag{9}$$

where $p_{treated,c}$ and $p_{control,c}$ are the proportions of category c in the treatment and control group, respectively and C is the set of categories a categorical variable can take. The results for the differences are shown in Table 2, the interpretation of the values is given in chapter 4.2.

By and large, the results for the difference reinforce the assumptions for the relevance of the control variables made in chapter 3.2.3. Only the facility size (proxied by the quantity of production waste in the reporting year), was highly similar for both groups. Hence, the propensity score was calculated based on deal year, the location, hazard, and industry. To deal with the multilevel categorical variables location and industry in propensity score matching, binary indicator variables were introduced via one-hot encoding. Thus, in the propensity score model the condition of Formula 6 was adapted as shown in Formula 10 with

$$\begin{gathered} X_{n'} = X_n \Longleftrightarrow X_{h,k,i,t} = X_{h',k',i',t'} \\ \text{with } P(PE_n = 1|X_n) \cong P(PE_{n'} = 1|X_{n'}) \\ \forall n \in N, \forall n' \in N \end{gathered} \tag{10}$$

Here, the PSM matches a pair of cases that have approximately the same probability P of receiving treatment conditional on their vectors of characteristics. In the PSM model used in this study, the propensity score was estimated using logistic regression and is used to match observations with a similar balance of control variables with a k-nearest neighbour algorithm. The matching was conducted using a Python programming environment.[75]

3.3.2. Deterministic matchig: Blocking and Matching

Due to the high relevance of categorical variables as confounders in this research, an adapted blocking and matching (BaM) method was developed and employed on the dataset.[76] Blocking is a method in big data analytics where records are grouped that share the same confounding variables.[77] In the case of the TRI dataset, first blocks based on exact record linkage of industry, state, hazard and year were created. Then for each treated facility in a block the closest untreated facility based on the quantity of production waste was found. The resulting pairs resemble tuples of a treated and an untreated facilities with similar (theoretically identical) control vectors. Thus for the blocking and matching the condition in Formula 6 was adapted as shown in Formula 11 with

$$\begin{gathered} X_{n'} = X_n \Longleftrightarrow X_{h,k,i,t,s} = X_{h',k',i',t',s'} \\ where\ n \overset{\text{def}}{=} n\ \forall n \in N, n' \in N \\ \Longleftrightarrow \left\{ h = h', k = k', i = i', t = t',\ s\ \approx s' \right\} \end{gathered} \tag{11}$$

The blocking approach for this research used deterministic blocking based on logical conjunction of year (deal year

[71] See Austin (2011b, pp. 399–424).

[72] See Cepeda et al. (2003, pp. 280–287).

[73] See Austin (2011b, pp. 410–411).

[74] See Austin (2010, p. 2140); The formula is adjusted by using a factor of 0.5 instead of 1/k, as the latter would give values too small for meaningful interpretation. Because the absolute difference in proportion was only needed to be informative for relative comparisons between groups or before and after matching, this adjustment was not detrimental.

[75] See Kline and Luo (2022, pp. 1354–1357); See appendix A p. 1 for the code used for PSM in this research.

[76] This was inspired by the generalized randomized block design according to Addelman (1969, p. 35) where blocking maximizes the covariance between treated and untreated samples based on their control variables, resulting in a minimum variance in the difference between treated and untreated samples. Essentially, this method aims to isolate any observed difference in treatment effect and attribute it solely to the effect of the treatment itself.

[77] See IBM Corporation (2021); Blocking and matching is commonly used in big data analysis to reduce computing power when examining data connections. Although the data structure of the TRI lends itself well to this methodological approach, it does not appear to have been used in research on the TRI.

Table 2: Balance of the covariates in treated and untreated group. a: Absolute quantities of toxic chemical in deal year in pounds.

Covariates	Difference
Continuous Variables$[d_{st}]$	
Year	0.4637
Total Releases[a]	-0.0314
Production Waste[a]	-0.0053
Categorical Variables $[d_{abs}]$	
Location	0.1446
Industry	0.2428
Chemical hazard	0.0175

= record year) and the categorical variables industry, state and hazard level. The blocking proceeded as follows: Two datasets were separated from the original database based on the status $PE = 1$ and $PE = 0$. Then, for each case in the $PE = 1$ sample, all cases from the $PE = 0$ sample were found via exact record linkage (i.e., with exactly the same combination) on year, industry, state and hazard level.[78] Within each block for the $PE = 1$ constituent the PE = 0 case which was closest on quantity of toxic chemical in production waste was found with the pandas *merge_asof* function to account for the proxied facility size.[79] The blocking and matching method proved to be a computationally intensive process, requiring approximately 2 hours to complete the matching.[80]

4. Empirical results

In the following, first the descriptive statistics of the final sample are shown. Then, the efficiency and quality of the matching methods are elaborated and lastly, the results of the matching for the estimated effect size of private equity takeover are presented.

4.1. Descriptive statistics

In Table 3 the summary statistics on the variables are shown. The control group is shown in Panel A and the treatment group in Panel B. It presents the number of observations, the mean, the standard deviation, the minimum, the 25 %, 50 %, 75 % quartiles and the maximum. From the final sample which consisted of 609,916 total and 1,746 deal-related facility-hazard-year observations, the dependent variables according to Formula 1 could be calcuated for approximatly two thirds of the cases (see column *N* in Table 3). As the dependent variables represent relative changes, these are highly sensitive to outliers in the underlaying data.[81] This manifests in the dataset by exaggeratedly great erroneous values (see column *Max* in Table 3) which significantly distort descriptive statistics (see the column *Mean* in Table 3). On the other hand, the minimum (see column *Min* in Table 3) of *-1* in the total releases indicates, that a facility completely stopped emitting any quantities associated with the chemical in the measured timeframe. Likewise, the minimum of *-1* in the production waste indicates that the facility no longer produces any waste containing the chemical. Especially for small facilities, it is also possible that the amount of the chemical handled has fallen below the reporting threshold. In addition to the presence of errors, high levels of disproportionality characterize quantities of toxic chemicals at the facility level in the TRI dataset.[82] Both, errors and disproportionalities suggest that the descriptive statistics fundamentally mischaracterize the environmental performance of facilities.[83] A skewness of 30 and a kurtosis of roughly 900 for the distribution of the dependent variables in Panel A, strongly disfavor the case of normality and reinforce this conjecture. To introduce robust statistical analysis which is less sensitive to the outliers, the median and the median absolute deviation (MAD) were used instead of the mean and the standard deviation.[84] To deal with outliers, the approach of Leys et al. (2013) was followed and subsequently cases greater than 2.5*MAD +- median were removed from the analysis.[85]

The trimming was done for the difference in total releases and production waste individually. The percentile value of the cut-off value was calculated as a measure for the proportion of data not being classified as outliers. A cut-off value of 2.6471 means that any observation inferring that a facility increased its pollution by more than 264.71% was excluded from the subsequent analysis. Table 4 shows the robust descriptive statistics for the dependent variables after the trimming step.[86] The median is again used as the better suited measure for ratios. The number of observations was reduced by the trimming, but more than 90 % of the data remained in the analysis. The values for skewness and kurtosis of below 2 are considered as acceptable to assume normal univariate

[78] See Fellegi and Sunter (1969, pp. 1183–1210) for a detailed explanation of record linkage.
[79] See Petrou (2017, pp. 338–386); See appendix A p. 3 for the code used for BaM in this research.
[80] The code was executed using parallelized threads according to Python Software Foundation (2023).
[81] See Miller (1993, pp. 457–459).
[82] See Collins et al. (2020, p. 2).
[83] See pp. 14-15 of appendix C for a graphical representation of the distribution of pollution in the TRI data.
[84] See Huber (2011, pp. 1248–1251).
[85] See Leys et al. (2013, pp. 764–766).
[86] See p. 20 in appendix C for graphical representation of the dependent variables of the trimmed sample.

Table 3: Summary of the descriptive statistics. This table summarizes the descriptive statistics of the unmatched data sample. Additionally, it presents the firm-specific and pollution-specific control variables. The unmatched sample is segmented according to whether a parent company of a facility was involved in a private equity takeover during a reporting year. a: Facility size is in quantities of toxic chemicals in pounds. * Statistics for the deal year other than Min and Max were omitted as not reasonably meaningful.

Panel A: Descriptive statistics PE = 0

Variable	N	Mean	SD	Min	25 %	Median	75 %	Max
Dependent Variables								
Difference Total Release	360,123	$6.44*10^8$	$2.28*10^{11}$	-1	-0.3931	-0.0422	0.3427	$9.98*10^{13}$
Difference Production Waste	389,902	$5.95*10^8$	$2.19*10^{11}$	-1	-0.3208	-0.0197	0.3211	$9.98*10^{13}$
Independent Variable								
PE	608,170	0	0	0	0	0	0	0
Firm-specific Control Variables								
Facility Location	608,170	-	-	-	-	-	-	-
Reporting Year	608,170	-*	-	1991	-	-	-	2021
Facility Size[a]	608,170	$1.11*10^6$	$1.79*10^7$	0	548.23	15,350	121,500	$3.75*10^9$
Industry	608,170	-	-	-	-	-	-	-
Pollution-specific Control Variable								
Environmental Hazard	608,170	2.0607	1.1516	1	1	2	3	4

Panel B: Descriptive statistics PE = 1

Variable	N	Mean	SD	Min	25 %	Median	75 %	Max
Dependent Variables								
Difference Total Release	1,142	35.66	711.45	-0.9999	-0.3584	-0.0194	0.3961	22,750
Difference Production Waste	1,241	25.47	659.74	-0.9939	-0.2825	-0.0148	0.3043	22,272
Independent Variable								
PE	1,746	1	0	1	1	1	1	1
Firm-specific Control Variables								
Facility Location	1,746	-	-	-	-	-	-	-
Reporting Year	1,746	-*	-	1991	-	-	-	2021
Facility Size[a]	1,746	$1.03*10^6$	$1.10*10^7$	0	984.87	20,345	117,932	$4.27*10^8$
Industry	1,746	-	-	-	-	-	-	-
Pollution-specific control Variable								
Environmental Hazard	1,746	2.0710	1.1489	1	1	2	3	4

distribution.[87] To summarize, eliminating outliers to obtain robust statistics improved data quality, but it also introduced an omitted measurement bias by excluding roughly 10 % of the data.

The median is slightly different between the treated and untreated groups but the differences should not be further interpretated because of the different number of observations. However, the difference for total releases and production waste in the PE = 0 sample can be taken as the median baseline reduction of quantities in any given reporting year, which is -8.6 % for total releases and -5 % for production waste. To summarize, while trimming effectively induced normality required for univariate data analysis, the differences are not suited to estimate impact effect of private equity takeover both on the change in quantities in total releases and production waste, respectively.

[87] A general convention is that a skewness less than 3 refers to a degree of symmetry of the normal distribution. See Burdenski (2002, p. 16) for further information.

4.2. Matching evaluation

For matching, from the original final sample with 609,916 facility-hazard-year observations, all cases with insufficient data for the difference in production waste data field were removed to reduce computing time, resulting in 391,143 observations subjected to the matching. The result of the matching is given in Table 5. PSM was able to match 100 % of the treatment group with facilities similar on their propensity logits from the control group since no caliper width was used. BaM was able to match 94 % of treated facilities to untreated facilities in the same year, state, and industry and on similar amount of production waste.

Before estimating the treatment effect, the matched sample is evaluated based on the quality of the matching. This assessment is critical to determine whether the imbalance of control variables has been adequately reduced, ensuring the reliability of the subsequent analysis. Thus, the standardized difference (Formula 8) and the absolute difference in proportions (Formula 9) were calculated for the matched sample. The results are shown in Table 6.

Table 4: Robust statistics for the dependant variables after trimming. Cut-Off at 2.5 x median absolute deviation (MAD) ± median (for trimming the cut-off is value is calculated based on the MAD and the median of the untrimmed distribution).

Variable	N	Median	MAD	Upper Cut-Off	Cut-Off Percentile	Skewness	Kurtosis
PE = 0							
Difference Total Release	331,719	-0.0863	0.3106	2.6471	92.11 %	1.3088	2.5973
Difference Production Waste	362.581	-0.0505	0.2742	2.4436	93.05 %	1.2181	2.4887
PE = 1							
Difference Total Release	1.029	-0.0760	0.2960	2.2837	90.11 %	1.0236	1.4907
Difference Production Waste	1.161	-0.0429	0.2480	2.3499	93.52 %	1.1996	2.1494

Table 5: Matching results. a: Each case presents a facility-hazard-year observation which has a value for the difference in for the production waste.

	PSM	BaM
Before matching		
Treated cases[a]	1,241	
Untreated cases[a]	391,143	
After matching		
Matched cases	1,241 (100 %)	1,169 (94.20 %)

Table 6: Standardized differences before and after matching.

	Unmatched	PSM	BaM
Continuous Variables $[d_{st}]$			
Year	0.4637	0.3419	-0.0207
Total Releases[a]	-0.0314	-0.1767	-0.0234
Production Waste[a]	-0.0053	0.0847	0.0005
Categorical Variables $[d_{abs}]$			
Location	0.1446	0.4214	0.0278
Industry	0.2428	0.4891	0.0306
Chemical hazard	0.0175	0.1708	0.0136
Imbalance reduction			
d_{st}	-	-20.56 %	91.09 %
d_{abs}	-	-167.05 %	82.21 %

The values for continuous variables before matching suggest that the groups exhibit considerable differences in terms of reporting years. The positive value indicates that treated sample possesses a higher average reporting year than the control sample, implying that private equity deals tend to be more prevalent in more recent reporting years.[88] The negative values for the quantities of production waste and total releases respectively indicate, that treated facilities have on average slightly lower amounts of toxic chemicals in those respective data fields. The values for the categorical variables show that, in the unmatched sample, the groups differ considerably in the proportions of facilities belonging to the respective industries. Although to a lesser extent, the distribution of facilities across states also differs between the groups. The groups show less variation in the levels of hazards present in the facilities. The propensity score matching provided a slightly better match rate and reduced imbalance for reporting year, but the value of 0.3419 still suggests high imbalance. PSM reduced imbalance between samples in terms of reporting years, but at the expense of imbalance in other control variables, ultimately resulting in 20 % and 167 % higher imbalance for continuous and categorical control variables, respectively. This is explained by the categorical nature of 3 of the 5 covariates, which necessitated their implementation through one-hot coding, resulting in a large number of binary covariates for which PSM has been shown to be generally weak to generate model dependence.[89] To summarize the values in Table 6 show that the PSM method was unsuccessful in achieving balance be-

[88] See p. 13 in appendix C for the number of reporting facilities per reporting year in the respective groups.

[89] See King and Nielsen (2019, pp. 435–454).

tween the treatment and control groups. Additionally, the findings support the strategy of treating the year as a categorical variable and subject it to exact record linkage matching, as opposed to considering it as a continuous variable in the propensity score calculation. This approach is especially crucial because comparisons across facilities may be significantly skewed by alterations in the TRI reporting framework in certain years, particularly in the case of lower reporting thresholds for specific chemical types.[90] Therefore, this study relied on blocking based on categorical covariates in the second matching method to ensure balance between treatment and control groups. Indeed, the BaM method successfully reduced the imbalance by over 80% across all control variables. Furthermore, all continuous control variables were balanced to a value for the difference of less than 0.1, which is generally considered to represent a negligible difference between the samples in the mean of control variables.[91]

It is crucial to note that the standardized difference and the absolute difference in proportions reflect the overall similarity between the two samples. Although these measures reveal improvements resulting from the matching methods employed, they do not allow an evaluation of the similarity of individual 1:1 matched pairs. To further assess the quality of the matching, the propensity scores as the probability of being assigned to one treatment, given an observation's measured baseline covariates was calculated for both matched samples. The resulting propensity score distribution is shown in Figure 3 for both methods.

The distribution in Figure 3a shows that both, the treated and control group exhibit similar propensity scores after the blocking and matching method. This indicates that the matching has been effective in balancing the covariates between the treatment and control groups. The peak at ~ 0.5 suggests that many of the matched samples have roughly equal probabilities of being in either the treatment or control group based on their covariates.

The Gaussian curves of the propensity scores depicted in Figure 3b for the treatment and control groups after the Propensity Score Matching reveal substantial non-overlapping distributions. This observation suggests that the propensity scores of a considerable number of samples differ, and matches have been paired that exhibit variations in their probability of receiving treatment based on their control variables. Overall, after the PSM the covariate distributions between the treatment and control groups still differ dramatically. A maximum caliper width was not used in this analysis as this would have resulted in a limited number of matches.[92]

4.3. Estimation of treatment effects

In order to test hypotheses 1a and 1b, which constitute that the impact of private equity on target companies manifests in a reduced amount of toxic chemicals emitted or wasted, for each matched pair, the effect size for production waste and total releases was calculated according to Formula 7. Since the effect size is no longer a ratio, trimming was not necessary for the resulting data and the median was used as an immanent robust measure. Hence the sample median treatment effect on the treated was estimated.[93] Table 7 summarizes results obtained for the effect of private equity ownership on the facilities. It compares the two matching methods employed and presents the number of pairs, the sample median treatment effect for the treated and the median absolute deviation. The unit of the effect size is in percentage points.

The estimated effect sizes from both methods are similar; however, the values derived from the blocking and matching approach are considered more reliable due to the poor quality of the propensity score matching. The value for the median effect size for total releases obtained through the BaM method suggests that the average quantities released at a facility two years after a private equity takeover of its parent company, exhibit a median increase of 1.5535 %-points relative to a similar facility that did not experience a change in ownership. Conversely, the quantity of production waste in the PE-backed facility is on median 1.1090 %-points lower following the private equity takeover of its parent company relative to non-acquired peers.

To differentiate between severity of the released and wasted quantities and to test hypotheses 2a and 2b, the matches were grouped by environmental hazard h to distinguish between hazard levels of toxic chemicals according to the developed environmental hazard level. The sample median treatment effect on the treated can thereby be observed for different hazard levels. The results are shown in Table 8.

When differentiating between hazard levels, the discrepancies between the matching methods become more apparent and as before, the values obtained from the blocking and matching approaches are considered more reliable. The number of observations varies between the hazard scores (see column N in Table 8), and a majority of chemicals used pose minimal hazards and releases are of low severity according to the environmental hazard model of this thesis. The number of observations for PBT-only substances (hazard score 2) is even lower than for carcinogen-only substances (hazard score 3). The median values for total releases of hazard score 2 and hazard score 4 quantities, with differences of 5.4541 and 11.4158 %-points, respectively, indicate that private equity-backed facilities significantly increase pollution with these chemicals after takeover. The situation is different for production waste. There, all hazard levels except hazard level 2 show a decrease in quantities compared to facilities that have not undergone a private equity takeover. Figure 4 provides a graphical representation that, although only informative, highlights the overarching patterns regarding the impact of private equity acquisitions contingent upon the environmental hazard. Interestingly, the trend direction for the

[90] See for example EPA (1999).

[91] See Normand et al. (2001, pp. 388–398).

[92] See Austin (2011a, pp. 151–161); The caliper width is the distance by which the propensity scores of a matched pair is allowed to differ by at most which avoids matching of highly dispersed propensity logits.

[93] Adapted from Imbens (2004, pp. 5–6).

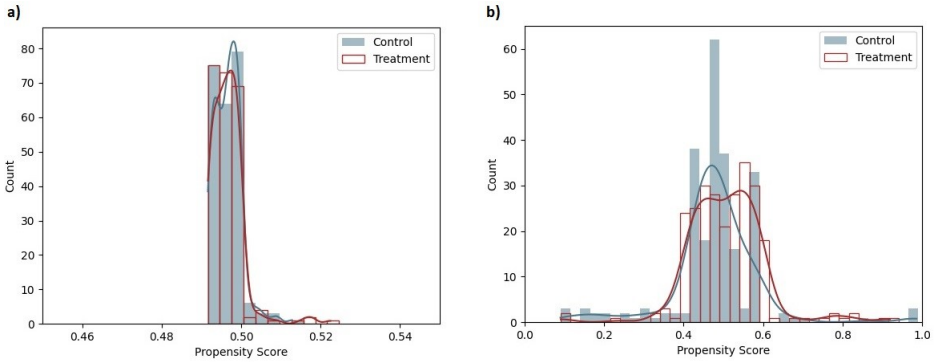


Figure 3: Calculated propensity score distribution of the matched samples. a) Matching based on BaM. b) Matching based on PSM. The two graphs are not to scale as the x-axis is zoomed in in Figure 3a.

Table 7: Estimated treatment effect on the treated. PSM stands for propensity score matching, BaM stands for blocking and matching. MAD stands for median absolute deviation from the median.

	Method: PSM			**Method: BaM**		
Variable	N	Median	MAD	N	Median	MAD
Effect size Total Release	1,227	1.0392	1.0444	1,080	1.5535	0.9168
Effect size Production Waste	1,241	-1.9657	1.0461	1,169	-1.1090	0.9746

Table 8: Estimated treatment effect on the treated distinguishing between hazard levels 1 to 4. PSM stands for propensity score matching, BaM stands for blocking and matching. MAD stands for median absolute deviation from the median.

	Method: PSM			**Method: BaM**		
Variable	N	Median	MAD	N	Median	MAD
Effect size Total Release						
Hazard level 1	596	-1.3920	0.0212	526	0.0130	0.0177
Hazard level 2	146	-4.9290	0.0646	126	5.4541	0.0572
Hazard level 3	314	-0.5266	0.0127	284	0.6299	0.0110
Hazard level 4	171	13.0623	0.1601	144	11.4158	0.1484
Effect size Production Waste						
Hazard level 1	602	-3.6206	0.0821	583	-0.9953	0.0217
Hazard level 2	147	1.9211	0.0215	131	4.0495	0.0482
Hazard level 3	316	-3.9517	0.0649	297	-1.9834	0.0430
Hazard level 4	176	1.9850	0.0393	158	-3.6917	0.0775

difference in production waste and total releases is diametrically opposed. As the hazard scores are derived from an ordinal scale, the ranking depends on the interpretation of the underlying severity of a hazard. In general, for both production waste and total releases, the effect size is greater for substances that are both carcinogenic and a PBT (hazard level 4) compared to substances that are either carcinogenic (hazard level 3) or PBTs (hazard level 2). Yet, private equity-backed facilities appear to be less concerned about substances that are PBTs (hazard level 2) compared to chemicals that are neither PBTs nor carcinogens (hazard level 1).

By and large, these trends suggest a hazard-dependent treatment effect for the treated facilities, where higher hazard is directly correlated with increased total releases and inversely correlated with decreased production waste.

5. Conclusion

The following section first draws the main conclusions from the findings. Then the chapter continues with a reflection on the limitations of the research imposed by the data and methodology. Finally, conjectures and ideas offer possible further research questions.

5.1. Discussion

This thesis investigates the evolution of environmental pollution emanating from facilities operated by companies that have experienced private equity takeovers in the United States between 1991 and 2021. It is one of the few academic papers to use the Toxic Release Inventory to assess ESG issues in the context of private equity and among those, it is

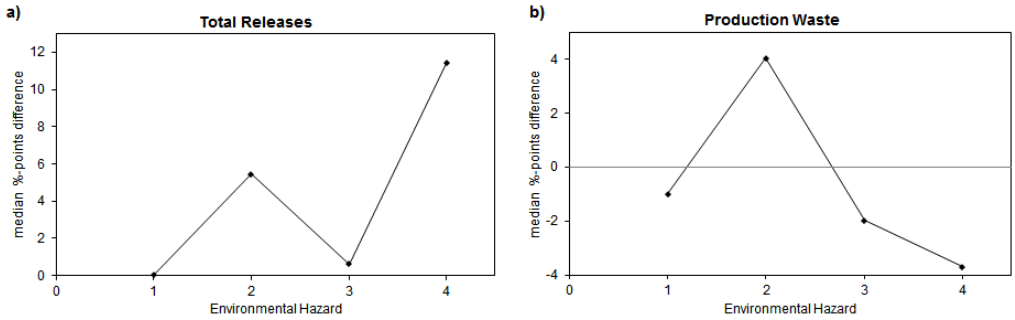


Figure 4: Graphical representation of the difference in %-points discriminating between environmental hazard level 1 to 4. The graphs show the difference in %-points obtained by the BaM method.

the first[94] to a) cover an extensive 30-year period, providing a comprehensive longitudinal perspective on the issue and b) utilize data on pollution and production waste as well as information about the environmental hazards of the chemicals thereof, thereby employing the TRI's extensive scope of information provided. The plethora of data available in the TRI allowed examining a period of two years before and after the deal for each facility. Therefore, this study effectively captures a five-year timeframe to evaluate the impact of private equity ownership on environmental performance which reduces influence of possible fluctuations in the environmental pollution due to one-time releases or other incidents. In addition, by excluding the environmental performance in the deal year itself from the analysis, the study avoids potential concerns of facilities deceiving sustainability by transiently reducing emissions in the deal year in order to pander private equity investors' ESG-due-diligence frameworks. For the research on asset-level sustainability two key aspects of the TRI-program append a unique value to the database: First, the database exhibits a comprehensive scope of data dating back to a time when ESG was far from established as an investment theme, thereby representing a remarkable standard of unconfoundness and data quality, which is ensured by capturing objective, granular information on various forms of releases through quantitative chemical analysis methods. Secondly, the TRI program only monitors, and does not impose restrictions on the use of chemicals. Although compulsory disclosure of information regarding the handling of toxic chemicals by industries to the public generates an incentive to enhance environmental performance,[95] this feature is crucial in the context of the research design in this thesis as the lack of legal pressure to reduce pollution from the program itself leaves the causality for the change in pollution level of a TRI-covered facility solely to the implication of universal factors, for example state laws and regulations as well as technological advances, organizational structure or both of the latter.[96] Setting aside such external effects, a main rationale in this thesis was, that when controlling for financial or judicial pressure, a decrease in pollution can be interpreted as a reflection of the investors (i.e., facilities owners) willingness to adopt sustainable practices. By using the TRI, this thesis also successfully overcomes the challenge of data scarcity, which is a pertinent problem in studying the private equity industry, especially when analyzing non-financial information about target firms. Yet, this challenge becomes apparent in the light of the sample construction: Out of 1.45 million observations in the TRI, less than 0.3% were attributable to firms involved in a private equity deal between 1991 and 2021. Half of these transactions in the datasample were buyouts according to Preqin and the majority took place after the year 2000.[97]

The methodological approach employed in this thesis utilizes robust methods to estimate the effect size of private equity takeovers and developes a novel matching method based on blocking and matching that significantly outperforms the alternative probabilistic matching based on propensity scores by up to over 200%. Matching is essential for studying the influence of private equity ownership on environmental performance, as it helps mitigate problems of selection bias arising from the tendency of private equity firms to favor certain industries. For instance, PBTs are chemicals that are typically used for very technology specific purposes, with electronics and computer products being a major application. In this industry, PE showed a positive selection bias, with 3.64 % of facilities in this sector in the full sample compared to 6.91 % of facilities in this sector in the PE=1 subset. The exact record linkage based on industry and hazard score in the BaM method used in this thesis avoided influence of such confounding factors. The novel matching process effectively reduces omitted variable bias that arises from facility- and pollution-fixed effects by comparing pairs of facilities that share both, idiosyncratic characteristics, and the environmental ramifications of their operations. This allowed controlling for facility size, industry, changes in productivity, as well as legal and other externalities which are due to location, environmental hazard of pollution and time. Additionally, the blocking and matching method developed in this study employes matching on similar quantities of production

[94] To the best of the author's knowledge.
[95] See for example Konar and Cohen (1997, pp. 109-124) or Saha and Mohr (2013, pp. 284-291).
[96] See for example Prechel and Zheng (2012, p. 950) or Mary L Streitwieser (1994, p. 2).
[97] For summary information on the transactions, see appendix C, p. 2.

waste, which addresses a potential selection bias exhibited by private equity investors who use ESG as a risk framework to select facilities that are already clean and do not require an investor to implement further pollution or waste reduction. As for the innovated blocking and matching approach greater accuracy and reliability in the analysis is ensured, the results obtained from BaM are considered for further interpretation.

The findings of this thesis reveal a general trend in which facilities reduce their pollution in terms of total releases by about 8.6 % per year and the amount of toxic chemicals as production waste by about 5.1 % per year over the period from 1991 to 2021.This baseline reduction aligns with observed trends in existing literature.[98] In the context of the impact of private equity on pollution, this study finds that the annual pollution reduction rate is 1.55 %-points lower if a facility's parent company is acquired by a private equity investor in the same year (i.e., a facility indirectly involved in a PE deal reduces pollution by about 7.05 % in the deal year). Conversely, the research finds that the reduction in production waste is 1.11 %-points higher in a year when a facility is indirectly involved in a PE deal (i.e., a facility reduces the amount of production waste by approximately 6.21 % in that year). In examining whether PE investors discriminate between different levels of environmental risk, the thesis further finds that higher chemical hazard is directly correlated with increased total releases and inversely correlated with decreased production waste. Conclusively, while the hypotheses theorized under the rationale of the private equity business model more effectively internalizing the social costs of pollution to maximize shareholder value suggest that private equity takeover leads to better overall pollution management in the target company, the implications drawn from the data of this thesis reveal a contrary image. First, it appears that the average private equity investor does not leverage its controlling stake in a company to encourage cleaner facilities and instead results in worse sustainability of the facilities. Thus, under the general view applied in the research of this thesis, the private equity business model clearly fails to internalize the social costs of environmental pollution. The findings infer that private equity investors instead have an incentive to avoid internalizing social costs of pollution, and this inclination becomes more salient in situations where associated costs are higher, i.e., in the case of pollution with a substance exhibiting high environmental hazard. Echoing the environmental hazard model interpretation, a general positive correlation of increased environmental pollution with increasing hazard is indeed observed. Interestingly, as PBTs persist in the environment for long periods of time and generally exhibit constant time-dependent degradation, they can be easily traced back to specific times and quantities of their releases. As a result, private equity investors may still remain liable even after divestiture of the asset, which is especially relevant for substances classified as environmental hazard class 4 (both PBT and carcinogenic), as they can cause severe damage long after their release and therefore have a high potential for lingering liability. Although this potential liability risk aligns with future social costs it is not internalized in the overall view of this research. The higher increase for substances with higher hazards may be attributed to the fact that pollution abatement measures for chemicals of high hazards are typically already using the best available technology (as typically required by law) and the increase in productivity after private equity takeover resulting in higher tonnage of toxic chemicals handled cannot be offset by these measures because they are already operating at their limits. This might elucidate the measurable increase despite controlling for changes in productivity and suggests a scenario where PE investors, yet aware of liability risks, still favor generating present profits over preparing for future penalties. This reality is supported by the work of Shive and Forster, who found significantly more judicial actions and higher penalties for PE-backed facilities compared to private independent backed facilities.[99]

The implications of production waste can be divided into two distinct, yet entangled narratives. The first narrative, the ESG-narrative, considers quantities of toxic chemicals in the production waste as risk factor for possible future occurrence of social costs which may arise from pollution or the imposition of Pigouvian taxes (see chapter 2.2). The other narrative considers quantities of toxic chemicals in production waste as an opportunity for value creation due to operational improvement, as waste represents wasted resources and financial losses. Since operational improvements at the target company level have been identified as a key driver of private equity sponsors' return,[100] in the quantity of toxic chemical in production waste, environmental and financial objectives are tied together where reduction of toxic (and expensive) chemicals results in cheaper (and cleaner) operations. As more hazardous substances are typically more expensive (both costs of the chemical itself and associated safety measures as well as allowances for the workers handling such substances), the incentive to reduce production waste increases with potential threat coming from the production waste. In the light of the results of this research, that a) the quantity of toxic chemical in production waste is reduced after private equity takeover and b) this reduction is more pronounced for substances that pose greater hazards, this finding reinforces the deduction that the private equity business model is only effective in achieving non-financial goals when these are well aligned with the financial ones.

Notwithstanding the first key finding of this thesis, which indicates that private equity ownership has a negative impact on the pollution levels of U.S. facilities, the theoretical foundations of the impact of capital deployed – which unequivocally show that a policy of engagement is superior to simply impact divesting[101] – highlight the realization for LPs pursuing socially responsible impact of their investments that the

[98] See for example Collins et al. (2020, p. 4).

[99] See Shive and Forster (2020, p. 1320).

[100] See Achleitner et al. (2011, pp. 155–156).

[101] See Berk and van Binsbergen (2021, p. 26).

private equity asset class is a viable instrument for achieving their objectives – if LPs reconcile this realization by the second key finding of this thesis that these objectives must be well aligned with incumbent financial objectives.

5.2. Limitations of the research

Databases such as the TRI are essential for measuring the relationship between economic activity and environmental pollution in the United States.[102] The employment of the TRI in academic research has led to various insights into the effects and the broader economic impact of pollution and polluting businesses, respectively.[103] More specific to the research on the effect of private equity ownership on environmental pollution, the TRI database has been used to uncover factors influencing asset-level sustainability, for example liability risks imposed by geographic location.[104] Nevertheless, it must be noted that the TRI database incurs several limitations with respect to the data availability for a single facility or a company holding such facilities. First, the TRI only refers to facility level sustainability and not portfolio company level sustainability. Besides that, another major impingement on the utilization of the data is the threshold for the quantity of substances at which a facility must begin monitoring the substance for a given calendar year. The ramification is, that facilities that are required to report are typically larger ones owned by publicly traded companies, while small- and medium-sized facilities (i.e., parent companies), which are typically targets of private equity deals in the US, are less extensively covered by the TRI database. In addition, a facility does not report for a calendar year if the amount of substance handled in that year is below the threshold, which hampers a year-to-year comparisons to a certain extent, but more so for small and medium-sized facilities. Another important caveat, especially for longitudinal use of TRI data, is that reporting standards and regulations (e.g., chemicals listed, thresholds, industries covered) have changed over time. This was especially true in the early years after 1987, and data prior to 1991 are generally considered to be of limited quality,[105] especially because the reporting threshold was 75.000 pounds in 1987, 50.000 pounds in 1988 and was set to 25.000 pounds in 1989.[106] For this reason, the panel data in this paper starts with the reporting year 1991 to analyze the pollution level of facilities. To facilitate data analysis, the EPA periodically updates the database from previous reporting years to reflect updates and mitigate disproportionality; however, the presence of minor inconsistencies in reporting standards must be considered as an inevitable limitation when interpreting the results of this thesis. Other limitations of the TRI are that it does not cover all industry sectors, not every facility within covered sectors is mandated to report to TRI, and from an environmental perspective, the TRI chemical list does not encompass all toxic chemicals utilized in the United States.

Regarding the environmental hazard model, the findings of this thesis infer a hazard-dependent treatment effect. It is important to note that this influence may be largely attributable to the ordinal hazard score developed in the study, which may exhibit significant model dependence. Specifically, the fluctuations observed in the trends may be a result of the interpretation inherent in the environmental hazard model created for this research. Hence, the implications from considering different hazard scores are only considered to be informative. Yet, disentangling nuanced trends with such intricate dependencies has not been previously explored in the literature.

In the methodological approach, the research in this thesis faced the pertinent limitations that are frequently incurred by the propensity score matching method. A major reason for the deficient performance of the propensity score model in this research is that the categorical variables *Industry* (23 industries) and *Location* (52 states) were implemented as binary variables via one-hot encoding which resulted in 75 binary variables in the matching process (additionally to the year, facility size and environmental hazard variables). Coarsened Exact Matching (CEM) has been shown to be superior to propensity score matching for samples with imbalances due to higher order interactions,[107] and the CEM approach has been used to study the environmental performance of companies in similar contexts.[108] However, the method could not be successfully applied in the setting of this thesis because data formatting issues in the TRI database. As an alternative to CEM, principal component analysis in the propensity score matching model was tested as this has been shown to be expedient in economic research when addressing a large set of confounders in PSM.[109] However, the approach did not succeed and subsequently, efforts were guided towards developing the novel blocking and matching method.

The blocking and matching method has the limitation that the algorithm used in this research performes 1:many matching. Since the preceding exact blocking on 4 variables created a large number of small blocks compromised of subsets of facilities within which matching was performed, 1:many matching was beneficial in this case as it ensured reduced bias due to missing matches. The preceding exact blocking also immanently limits overestimation of the treatment effect due to multiple pairing of a control case, as it constraints the maximum number of control cases which can be matched to a treatment case. It must be noted however,

[102] See Bradley C. Karkkainen (2019).

[103] For a comprehensive review on research including the TRI, see Young et al. (2022).

[104] See Bellon (2020, p. 2).

[105] See Scott de Marchi and James T. Hamilton (2006, pp. 60–61).

[106] See "Toxic Chemical Release Reporting: Community Right-To-Know - PART 372" (1988)

[107] For a detailed elaboration on the advantages of CEM over traditional matching methods, especially propensity score matching, the reader should refer to Iacus et al. (2012, p. 2).

[108] For research using CEM as a matching method in the context of environmental performance in the US, see Hora and Subramanian (2019, p. 6), and in Europe see Kube et al. (2019, pp. 104558–104570).

[109] See for example Griffin et al. (2020, pp. 5537–5549).

that in case of less stringent blocking, greedy 1:1 matching should be performed.

5.3. Further research

The findings of this thesis offer intriguing research questions that could reveal relevant ramifications for policymakers as well as for market participants in the private equity industry. As elaborated above, ESG-considerations can protect future revenues from exposure to threats when these eventually become substantial. Thus ESG-performance equivalates possible financial outperformance in a more distant future. According to the results of this thesis, this possible financial return is not harvested through internalization of present social costs related to environmental pollution. In this context, current research revealed that the timeframes of asset managers are often too short to evaluate ESG-performance of assets.[110] Hence, further research could focus on determining whether a) private equity firms exhibit inability to accurately evaluate ESG risks and thus fail to address relevant topics at the portfolio company level or b) the private equity business model exhibits an investment period that is too short to internalize long-term social costs. To address this question, the overall impact of private equity, as measured in this thesis, could be disaggregated by investor characteristics to identify factors that foster an ESG-friendly phenotype among PE investors. In this context, Abraham et al. (2022) discovered that PE investors who emphasize ESG as a value driver on their website contribute to a decrease in pollution after takeover, relative to investors who do not adhere to SRI principles.[111] In addition, the research could explore the "skill or luck" question[112] to determine whether the ESG performance of these investors is due to their ability to drive sustainable change or merely the result of serendipity.

In the context of the relative decrease of the amount of production waste, it would be intriguing to investigate whether the reduction is a directly aspired outcome of the operational engineering that private equity firms undertake to enhance efficiency after takeover, or attributable to the alleviation of financial constraints that enable technology investment and thereby indirectly reduce hazard potentiality as a byproduct. To address this question, hypothetical differences between facilities supported by industry-agnostic PE investors and those with industry expertise could be explored. Both of the above research questions have important implications for policymakers in determining how to effectively regulate the private equity business model to position it as a catalyst for transforming the manufacturing sector toward better sustainability, while ensuring that private equity continues to play a constructive role in delivering financial value to limited partners.

Echoing the hazard score interpretation, the model in this thesis relied on an ordinal scale and exhibited significant model dependance. This might explain, why private equity takeover appears to induce an increase in the total releases and production waste of substances classified as PBTs (environmental hazard level 2) compared to substances classified as neither PBT nor carcinogen (environmental hazard level 3). While for the production waste this increase might be due to aforementioned model dependance, the general positive correlation of increased environmental pollution with increasing hazard suggests a tendency of PE investors to care less about the environment. Further research is needed to distinguish between hazard scores and could for example employ the full RSEI model from the EPA.

Along these lines, a potential further research question could focus on deal characteristics and how they relate to ESG-performance of target facilities. Arguably, certain deal types, such as those involving LPs investing through funds of funds, could serve to mask investments in environmentally detrimental assets. Such investments could then be used to offset lower returns from environmentally sustainable ventures. This thesis did not delve into the intricacies of deal structures; therefore, future studies could focus on addressing the limitations of missing data from the Preqin data source and further investigate the potential impact of deal characteristics on ESG performance.

Lastly, to engender a holistic view on the ESG performance of the PE-backed facilities examined in this thesis, the S and the G pillar are eminent topics of further research. In this thesis the focus is on the “E” (i.e., environmental) component of ESG and it would be interesting to reveal the S pillar of ESG in the same context to give a detailed picture.[113]

[110] See Eccles et al. (2017, p. 128).

[111] See Abraham et al. (2022, p. 13).

[112] See Korteweg and Sorensen (2017, pp. 535–562) for the question of skill and luck in private equity performance.

[113] For example the National Establishment Time-Series database would provide the S component and can be used well together with the data from the TRI (Technische Universität München (2023)).

References

42 U.S.C. 11023 and 11048. (1988). Toxic Chemical Release Reporting: Community Right-To-Know - PART 372. Retrieved February 12, 2023, from https://www.ecfr.gov/current/title-40/chapter-I/subchapter-J/part-372#372.23

Abraham, J. K., Olbert, M., & Vasvari, F. P. (2022). ESG Disclosures in the Private Equity Industry. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4265171

Achleitner, A.-K., Braun, R., & Engel, N. (2011). Value creation and pricing in buyouts: Empirical evidence from Europe and North America. *Review of Financial Economics*, *20*(4), 146–161. https://doi.org/10.1016/j.rfe.2011.09.001

Addelman, S. (1969). The Generalized Randomized Block Design. *The American Statistician*, *23*(4), 35–36. https://doi.org/10.2307/2681737

Agrawal, A., & Tambe, P. (2016). Private Equity and Workers' Career Paths: The Role of Technological Change. *The Review of Financial Studies*, *29*(9), 2455–2489. https://doi.org/10.1093/rfs/hhw025

Antoni, M., Maug, E., & Obernberger, S. (2019). Private equity and human capital risk. *Journal of Financial Economics*, *133*(3), 634–657. https://doi.org/10.1016/j.jfineco.2019.04.010

Atul Gupta, Sabrina T Howell, Constantine Yannelis, & Abhinav Gupta. (2021). Does Private Equity Investment in Healthcare Benefit Patients? Evidence from Nursing Homes. https://ideas.repec.org/p/bfi/wpaper/2021-20.html

Austin, P. C. (2010). The performance of different propensity-score methods for estimating differences in proportions (risk differences or absolute risk reductions) in observational studies. *Statistics in medicine*, *29*(20), 2137–2148. https://doi.org/10.1002/sim.3854

Austin, P. C. (2011a). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate behavioral research*, *46*(3), 399–424. https://doi.org/10.1080/00273171.2011.568786

Austin, P. C. (2011b). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical statistics*, *10*(2), 150–161. https://doi.org/10.1002/pst.433

Bauer, R., Ruof, T., & Smeets, P. (2021). Get Real! Individuals Prefer More Sustainable Investments. *The Review of Financial Studies*, *34*(8), 3976–4043. https://doi.org/10.1093/rfs/hhab037

Bellon, A. (2020). Does Private Equity Ownership Make Firms Cleaner? The Role Of Environmental Liability Risks. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3604360

Berg, F., Kölbel, J. F., & Rigobon, R. (2022). Aggregate Confusion: The Divergence of ESG Ratings*. *Review of Finance*, *26*(6), 1315–1344. https://doi.org/10.1093/rof/rfac033

Berk, J. B., & van Binsbergen, J. H. (2021). The Impact of Impact Investing. https://doi.org/10.2139/ssrn.3909166

Bernstein, S., & Sheen, A. (2016). The Operational Consequences of Private Equity Buyouts: Evidence from the Restaurant Industry. *The Review of Financial Studies*, *29*(9), 2387–2418. https://doi.org/10.1093/rfs/hhw037

Bian, Y., Gao, H., Wang, R., & Xiong, X. (2022). Sustainable development for private equity: Integrating environment, social, and governance factors into partnership valuation, 1–12. https://doi.org/10.1002/bse.3304

Bradley C. Karkkainen. (2019). Information as Environmental Regulation: TRI and Performance Benchmarking, Precursor to a New Paradigm? In P. S. Menell (Ed.), *Environmental law* (pp. 185–299). Routledge.

Burdenski, T. (2002). Evaluating Univariate, Bivariate, and Multivariate Normality Using Graphical Procedures. *Mult. Lin. Regression Viewpoints*, *26*(2), 15–28. Retrieved March 19, 2023, from https://imaging.mrc-cbu.cam.ac.uk/statswiki/FAQ/Simon?action=AttachFile&do=get&target=mahalplot.pdf

Bureau van Dijk Electronic Publishing Ltd. (2023). Our products: Orbis (Bureau van Dijk Electronic Publishing Ltd, Ed.). Retrieved February 13, 2023, from https://www.bvdinfo.com/en-gb/our-products/data/international/orbis

Cepeda, M. S., Boston, R., Farrar, J. T., & Strom, B. L. (2003). Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. *American journal of epidemiology*, *158*(3), 280–287. https://doi.org/10.1093/aje/kwg115

Cohn, J. B., Hotchkiss, E. S., & Towery, E. M. (2022). Sources of Value Creation in Private Equity Buyouts of Private Firms. *Review of Finance*, *26*(2), 257–285. https://doi.org/10.1093/rof/rfac005

Cohn, J. B., Mills, L. F., & Towery, E. M. (2014). The evolution of capital structure and operating performance after leveraged buyouts: Evidence from U.S. corporate tax returns. *Journal of Financial Economics*, *111*(2), 469–494. https://doi.org/10.1016/j.jfineco.2013.11.007

Cohn, J. B., Nestoriak, N., Wardlaw, M., & Jiang, W. (2021). Private Equity Buyouts and Workplace Safety. *Review of Financial Studies*, *34*(10), 4832–4875. https://ideas.repec.org/a/oup/rfinst/v34y2021i10p4832-4875..html

Collins, M., Pulver, S., Hill, D., & Manski, B. (2020). Characterizing disproportionality in facility-level toxic releases in US manufacturing, 1998–2012. *Environmental Research Letters*, *15*(6), 064002. https://doi.org/10.1088/1748-9326/ab7393

Crifo, P., & Forget, V. D. (2013). Think Global, Invest Responsible: Why the Private Equity Industry Goes Green. *Journal of Business Ethics*, *116*(1), 21–48. https://doi.org/10.1007/s10551-012-1443-y

Eaton, C., Howell, S. T., & Yannelis, C. (2020). When Investor Incentives and Consumer Interests Diverge: Private Equity in Higher Education. *The Review of Financial Studies*, *33*(9), 4024–4060. https://doi.org/10.1093/rfs/hhz129

Eccles, R. G., Kastrapeli, M. D., & Potter, S. J. (2017). How to Integrate ESG into Investment Decision-Making: Results of a Global Survey of Institutional Investors. *Journal of Applied Corporate Finance*, *29*(4), 125–133. https://doi.org/10.1111/jacf.12267

Elton, E. J., & Gruber, M. J. (1997). Modern portfolio theory, 1950 to date. *Journal of Banking & Finance*, *21*(11-12), 1743–1759. https://doi.org/10.1016/S0378-4266(97)00048-4

EPA. (1999). Persistent Bioaccumulative Toxic (PBT) Chemicals; Lowering of Reporting Thresholds for Certain PBT Chemicals; Addition of Certain PBT Chemicals; Community Right-to-Know Toxic Chemical Reporting: 40 CFR Part 372. Retrieved February 24, 2023, from https://www.govinfo.gov/content/pkg/FR-1999-10-29/pdf/99-28169.pdf

EPA. (2022a). Persistent Bioaccumulative Toxic (PBT) Chemicals Covered by the TRI Program. Retrieved April 15, 2023, from https://www.epa.gov/toxics-release-inventory-tri-program/persistent-bioaccumulative-toxic-pbt-chemicals-covered-tri

EPA. (2022b). Risk-Screening Environmental Indicators (RSEI) Model. Retrieved February 12, 2023, from https://www.epa.gov/rsei/rsei-data-and-modeling

EPA. (2022c). Toxic Chemical Release Inventory Reporting Forms and Instructions: Revised 2022 Version. https://ordspub.epa.gov/ords/guideme_ext/f?p=guideme:rfi:::::rfi:rfi

EPA. (2022d). TRI-listed toxic chemicals and chemical categories. Retrieved February 12, 2023, from https://ordspub.epa.gov/ords/guideme_ext/f?p=guideme:chemical-list-basic-search

EPA. (2023). What is the Toxics Release Inventory? Retrieved April 15, 2023, from https://www.epa.gov/toxics-release-inventory-tri-program/what-toxics-release-inventory

EUROSIF. (2019). The Sustainable Finance Disclosure Regulation (SFDR). Retrieved March 6, 2023, from https://www.eurosif.org/policies/sfdr/

Fauver, L., & Fuerst, M. E. (2006). Does good corporate governance include employee representation? Evidence from German corporate boards. *Journal of Financial Economics*, *82*(3), 673–710. https://doi.org/10.1016/j.jfineco.2005.10.005

Fellegi, I. P., & Sunter, A. B. (1969). A Theory for Record Linkage. *Journal of the American Statistical Association*, *328*(64), 1183–1210. https://doi.org/10.1080/01621459.1969.10501049

Franklin, B. A. (1985). TOXIC CLOUD LEAKS AT CARBIDE PLANT IN WEST VIRGINIA. *New York Times*, 1.

Friedman, M. (2007). The Social Responsibility of Business Is to Increase Its Profits. In W. C. Zimmerli, M. Holzinger, & K. Richter (Eds.), *Corporate Ethics and Corporate Governance* (pp. 173–178). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-70818-6{\textunderscore}14

Galbreath, J. (2009). Building corporate social responsibility into strategy. *European Business Review*, *21*(2), 109–127. https://doi.org/10.1108/09555340910940123

Gillan, S. L., Koch, A., & Starks, L. T. (2021). Firms and social responsibility: A review of ESG and CSR research in corporate finance. *Journal of Corporate Finance*, *66*, 101889. https://doi.org/10.1016/j.jcorpfin.2021.101889

Goodland, R. (1995). THE CONCEPT OF ENVIRONMENTAL SUSTAINABILITY. *Annual Review of Ecology and Systematics*, *26*(1), 1–24. https://doi.org/10.1146/annurev.es.26.110195.000245

Griffin, B., Hartarska, V., & Nadolnyak, D. (2020). Credit Constraints and Beginning Farmers' Production in the U.S.: Evidence from Propensity Score Matching with Principal Component Clustering. *Sustainability*, *12*(14), 5537–5549. https://doi.org/10.3390/su12145537

Heal, G. (2005). Corporate Social Responsibility: An Economic and Financial Framework. *The Geneva Papers on Risk and Insurance - Issues and Practice*, *30*(3), 387–409. https://doi.org/10.1057/palgrave.gpp.2510037

Hora, M., & Subramanian, R. (2019). Relationship between Positive Environmental Disclosures and Environmental Performance: An Empirical Investigation of the Greenwashing Sin of the Hidden Trade–

off. *Journal of Industrial Ecology*, *23*(4), 855–868. https://doi.org/10.1111/jiec.12823

Huber, P. J. (2011). Robust Statistics. In M. Lovric (Ed.), *International Encyclopedia of Statistical Science* (pp. 1248–1251). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-04898-2{\textunderscore}594

Iacus, S. M., King, G., & Porro, G. (2012). Causal Inference without Balance Checking: Coarsened Exact Matching. *Political Analysis*, *20*(1), 1–24. https://doi.org/10.1093/pan/mpr013

IBM Corporation. (2021). Blocking and matching concepts (IBM Corporation, Ed.). Retrieved March 30, 2023, from https://www.ibm.com/docs/en/iis/8.1?topic=jobs-blocking-matching-concepts

Imbens, G. W. (2004). Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review. *Review of Economics and Statistics*, *86*(1), 4–29. https://doi.org/10.1162/003465304323023651

Indahl, R., & Jacobsen, H. G. (2019). Private Equity 4.0: Using ESG to Create More Value with Less Risk. *Journal of Applied Corporate Finance*, *31*(2), 34–41. https://doi.org/10.1111/jacf.12344

Jochen Runde. (1998). Clarifying Frank Knight's discussion of the meaning of risk and uncertainty. *Cambridge Journal of Economics*, *22*(5), 539–546. Retrieved March 8, 2023, from http://www.jstor.org/stable/23600452

Kaplan, S. N. (1989). Management Buyouts: Evidence on Taxes as a Source of Value. *The Journal of Finance*, *44*(3), 611–632. https://doi.org/10.1111/j.1540-6261.1989.tb04381.x

Kaplan, S. N., & Strömberg, P. (2009). Leveraged Buyouts and Private Equity. *Journal of Economic Perspectives*, *23*(1), 121–146. https://doi.org/10.1257/jep.23.1.121

Kim, T., & Xu, Q. (2022). Financial Constraints and Corporate Environmental Policies. *The Review of Financial Studies*, *35*(2), 576–635. https://doi.org/10.1093/rfs/hhab056

King, G., & Nielsen, R. (2019). Why Propensity Scores Should Not Be Used for Matching. *Political Analysis*, *27*(4), 435–454. https://doi.org/10.1017/pan.2019.11

Kline, A., & Luo, Y. (2022). PsmPy: A Package for Retrospective Cohort Matching in Python. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, *2022*, 1354–1357. https://doi.org/10.1109/EMBC48229.2022.9871333

Knight, F. H. (1921). *Risk, Uncertainty and Profit*. Houghton Mifflin Company.

Konar, S., & Cohen, M. A. (1997). Information As Regulation: The Effect of Community Right to Know Laws on Toxic Emissions. *Journal of Environmental Economics and Management*, *32*(1), 109–124. https://doi.org/10.1006/jeem.1996.0955

Korteweg, A., & Sorensen, M. (2017). Skill and luck in private equity performance. *Journal of Financial Economics*, *124*(3), 535–562. https://doi.org/10.1016/j.jfineco.2017.03.006

Kube, R., von Graevenitz, K., Löschel, A., & Massier, P. (2019). Do voluntary environmental programs reduce emissions? EMAS in the German manufacturing sector. *Energy Economics*, *84*, 104558. https://doi.org/10.1016/j.eneco.2019.104558

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Liang, H., Sun, L., & Teo, M. (2022). Responsible Hedge Funds. *Review of Finance*, *26*(6), 1585–1633. https://doi.org/10.1093/rof/rfac028

Littig, B., & Griessler, E. (2005). Social sustainability: a catchword between political pragmatism and social theory. *International Journal of Sustainable Development*, *8*(1/2), 65–79. https://doi.org/10.1504/IJSD.2005.007375

Mary L Streitwieser. (1994). Cross Sectional Variation In Toxic Waste Releases From The U.S. Chemical Industry: Working Papers. https://ideas.repec.org/p/cen/wpaper/94-8.html

Mastromonaco, R. (2015). Do environmental right-to-know laws affect markets? Capitalization of information in the toxic release inventory. *Journal of Environmental Economics and Management*, *71*, 54–70. https://doi.org/10.1016/j.jeem.2015.02.004

Miller, J. N. (1993). Tutorial review—Outliers in experimental data and their treatment. *The Analyst*, *118*(5), 455–461. https://doi.org/10.1039/an9931800455

Normand, S. T., Landrum, M. B., Guadagnoli, E., Ayanian, J. Z., Ryan, T. J., Cleary, P. D., & McNeil, B. J. (2001). Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: a matched analysis using propensity scores. *Journal of Clinical Epidemiology*, *54*(4), 387–398. https://doi.org/10.1016/S0895-4356(00)00321-8

Olsson, M., & Tåg, J. (2017). Private Equity, Layoffs, and Job Polarization. *Journal of Labor Economics*, *35*(3), 697–754. https://doi.org/10.1086/690712

Petrou, T. (2017). *Pandas cookbook: Recipes for scientific computing, time series analysis and data visualization using Python*. Packt Publishing.

Prechel, H., & Zheng, L. (2012). Corporate Characteristics, Political Embeddedness and Environmental Pollution by Large U.S. Corporations. *Social Forces*, *90*(3), 947–970. http://www.jstor.org/stable/41682684

Preqin. (2023). Our Data. Retrieved February 6, 2023, from https://www.preqin.com/data/our-data

Python Software Foundation. (2023). concurrent.futures — Launching parallel tasks (Python Software Foundation, Ed.). Retrieved April 14, 2023, from https://docs.python.org/3/library/concurrent.futures.html

Renneboog, L., ter Horst, J., & Zhang, C. (2008). Socially responsible investments: Institutional aspects, performance, and investor behavior. *Journal of Banking & Finance*, *32*(9), 1723–1742. https://doi.org/10.1016/j.jbankfin.2007.12.039

Saha, S., & Mohr, R. D. (2013). Media attention and the Toxics Release Inventory. *Ecological Economics*, *93*, 284–291. https://doi.org/10.1016/j.ecolecon.2013.05.008

Scott de Marchi & James T. Hamilton. (2006). Assessing the accuracy of self-reported data: an evaluation of the toxics release inventory. *Journal of Risk and Uncertainty*, *32*(1), 57–76. Retrieved April 16, 2023, from http://www.jstor.org/stable/41761223

Shive, S. A., & Forster, M. M. (2020). Corporate Governance and Pollution Externalities of Public and Private Firms. *The Review of Financial Studies*, *33*(3), 1296–1330. https://doi.org/10.1093/rfs/hhz079

Shleifer, A., & Summers, L. H. (1988). Breach of trust in hostile takeovers. In Alan J. Auerbach (Ed.), *Corporate takeovers: Causes and consequences* (pp. 33–68). University of Chicago Press. http://www.nber.org/chapters/c2052

Shleifer, A., & Vishny, R. W. (1986). Large Shareholders and Corporate Control. *Journal of Political Economy*, *94*(3), 461–488. https://www.jstor.org/stable/1833044

Sørensen, M., & Yasuda, A. (2022). Impact of Private Equity. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4087778

Technische Universität München. (2023). National Establishment Time-Series (NETS). Retrieved April 9, 2023, from https://www.ub.tum.de/en/datenbanken/details/104981

UN PRI. (2019). About the PRI. Retrieved March 6, 2023, from https://www.unpri.org/about-us/about-the-pri

United Nations. (2023). The seventeen Sustainable Development Goals: Agenda 2030. Retrieved March 6, 2023, from https://unric.org/de/17ziele/

United States Census Bureau. (2022). North American Industry Classification System. Retrieved February 12, 2023, from https://www.census.gov/naics/

Young, B., Ingwersen, W. W., Bergmann, M., Hernandez-Betancur, J. D., Ghosh, T., Bell, E., & Cashman, S. (2022). A System for Standardizing and Combining U.S. Environmental Protection Agency Emissions and Waste Inventory Data. *Applied sciences (Basel, Switzerland)*, *12*(7), 3447–3463. https://doi.org/10.3390/app12073447

Zaccone, M. C., & Pedrini, M. (2020). ESG Factor Integration into Private Equity. *Sustainability*, *12*(14), 5725. https://doi.org/10.3390/su12145725

Junior Management Science 9(1) (2024) 1123-1139

Junior Management Science

www.jums.academy
ISSN: 2942-1861

How Do Companies Communicate Sustainability: A Semantic Analysis of German Automotive Manufacturers

Fabienne Le

Ludwig Maximilian University of Munich

Abstract

CSR is increasingly becoming an integral part of a company´s business operation. To successfully implement a CSR strategy, companies must address their CSR actions to their stakeholders. This paper examines how companies communicate sustainability to their stakeholders through various communication channels. This paper supports the novel strand of research applying computer-aided quantitative analysis methods as an alternative to qualitative methods, commonly used in business ethics and sustainability research. With the application of a latent semantic analysis, four automotive companies were examined for their sustainability communication. The paper offers new insights into the use of different communication channels, highlighting that companies address specific aspects of their CSR actions depending on what stakeholder group they want to address.

Keywords: communication channels; computer-aided quantitative analysis; CSR; latent semantic analysis; sustainability; triple bottom line

1. Introduction

Since the early 1990s, researchers and economists have called upon a new way of conducting business. Leaders should pursue to create not only economic value but also environmental and social value, therefore, making sustainability a main goal for businesses (Beal et al., 2017). Nowadays, sustainable development is considered integral to business operations (Schaltegger & Burritt, 2018). This shift has been attributed to two factors by prior research studies. For one, economic forces drive companies to increase profitability and improve their brand value. By implementing voluntary social and environmental activities, in form of corporate social responsibility actions (CSR), into a company's business strategy, companies can reduce manufacturing costs, lower operating risks, and build trust and customer loyalty (Kang et al., 2016; Russo-Spena et al., 2018). Moreover, stakeholders such as regulators, customers, and shareholders are becoming increasingly more attuned to sustainability topics. And further, are placing more pressure on companies to address challenges that come with it such as climate change and economic inclusion (Beal et al., 2017; Marcelino-Sádaba et al., 2015). Hence, stakeholder management and communication play a crucial role for a successful implementation of sustainable development strategies. By communicating their commitment and actions in a transparent and truthful manner, companies can mitigate legitimacy threats and establish a clear brand identity.

As researchers and companies recognize the importance of sustainability communication, multiple studies have been conducted on CSR disclosures and their impact on financial performance (Du et al., 2010; Miller et al., 2020), as well as the motivation behind sustainable development actions (Russo-Spena et al., 2018). Although there has been research conducted on CSR communication, most analyze one single channel, e.g., CSR reporting (Du et al., 2010; Mann et al.,

First, I want to thank Prof. Dr. Jelena Spanjol for giving me the opportunity to write my bachelor thesis, at the Chair for Innovation management. More so, I want to thank my supervisor Dr. Marcelo F. de la Cruz for his support and guidance. Lastly, I want to express my gratitude for Yasmin Le, Ella Ritchie and Lisa Kanis who always supported and encouraged my work.

DOI: https://doi.org/10.5282/jums/v9i1pp1123-1139

2021) or social media (K. Lee et al., 2013), while very few include both channels. Moreover, hardly any apply quantitative methods. Lock and Seele (2016) point out that there is a need to apply quantitative content analysis to the field of business ethics and CSR. This is because these fields traditionally use human-encoding and judgmental text interpretations which are more exposed to bias. Further, the context can only be represented on a one-dimensional level. Besides that, less than a handful of studies (Kountouri et al., 2019; Liao et al., 2018; Lock & Seele, 2016) include a computer-aided textual analysis.

However, the three previously mentioned studies focus on just one communication channel. This thesis addresses that gap by applying a latent semantic analysis (LSA), a computer-aided methodology for textual analysis, to assess how companies use different channels to communicate sustainability to different stakeholders. The use of LSA allows statistical text theme extraction while minimizing the risk of bias judgement. Through LSA but more so because this study emphasizes three channels, namely financial and non-financial disclosures (NFD), and social media, this thesis provides new insights about sustainability communication for multiple stakeholders. This thesis assesses four automotive companies on how they communicate their sustainable development actions to various stakeholders (shareholders, customers, and legislation) and compares them to each other.

The research offers new insights into the sustainability communication of companies. The findings indicate that companies emphasize different aspects of the triple bottom line framework (Elkington, 1998) when addressing specific shareholders. More so, the results highlight that companies must find a way to communicate all three aspects of the TBL in each channel. Selective disclosure as it has been observed in this research could lead to stakeholders perceiving the companies' communication as a greenwashing strategy. This could have grave implications for their brands and lead to reputational damages.

Furthermore, contributes this study to prior research that apply quantitative methods to the fields of business ethics and sustainability. It highlights that computer-aided quantitative textual analyses can be successfully applied to these fields of research and can be used as an alternative to traditional human-encoded methods which are more prone to bias interpretations of data.

The research paper is structured as follows: after outlining important aspects and frameworks behind sustainable developments and sustainability communication, the methodology section will provide specifics about the sample collection. Additionally, the data analysis process will be explained as well as how the outputs of LSA were interpreted. Following, the results will be discussed and interpreted. Lastly, the conclusion present theoretical, practical and policy implications together with limitations of the study and an outlook for future research.

2. Theoretical Background

2.1. Firms, Sustainability and Business Performance

2.1.1. The Triple Bottom Line

Sustainability is described as development that meets today's needs while preserving the future of coming generations and their chance to meet their own needs (United Nations, 2015). The triple bottom line (TBL) or 3Ps (People, Planet, Profit) is a framework coined by Elkington (1998) that was designed to include social and environmental dimensions into the previously finance-focused metrics of business performance (Elkington, 1998; Loviscek, 2020). It thereby indicates that companies must conform to societal expectations, minimize their negative environmental impacts, and maintain economic profitability. It is paramount for companies to balance all three pillars and pursue all three goals simultaneously to truly achieve sustainable development (Elkington, 1998; Lock & Araujo, 2020; Mish & Scammon, 2010; Sanchez-Chaparro et al., 2022). In terms of the social dimension, companies must build social capital by establishing long-term relationships with various stakeholders (Mann et al., 2021). Social capital entails improving lives and the community through fair treatment, implementing welfare policies, and providing educational opportunities for stakeholders and society. To further environmental goals, companies must secure as well as expand natural capital by not endangering the planet and minimizing the impact of its business operations on the environment. For instance, this could be in form of protecting land, air and water and using sustainable materials, products, or renewable energy. Lastly, goals of the economic dimension include the overall economic growth of the company and society, improving brand image and sales along with reducing production costs (Elkington, 1998; Mann et al., 2021). Although it is imperative to strike a balance between the three dimensions, they can be conflicting since stakeholders place varying weight on priorities at times (Goh et al., 2020; Mann et al., 2021). Past research shows that companies occasionally neglect environmental and societal aspects for the sake of economic opportunities because the financial outcomes can be better quantified and to maintain the company's positive, short-term cashflows (Elkington, 2018). To overcome frictions, companies must embrace and even take advantage of them by creating synergies that could enable sustainable development. These synergies and more importantly sustainable development can be achieved by establishing long-term partnerships, knowledge sharing, and conjointly founded solutions by private and public sectors, groups of main interest, and companies along the supply chain (Goh et al., 2020; Loviscek, 2020; Mann et al., 2021). This highlights that the TBL is a qualified framework to drive positive change as it evaluates social, environmental, and economic impacts across various stakeholders, e.g., community and legislation (Mann et al., 2021; Shinkle & Spencer, 2012).

2.1.2. Corporate Social Responsibility

When companies engage in corporate social (CS) activities they comply with societal rules, statutory regulations, and policies concerning social issue e.g., human rights or the environment. Notably, these policies only require companies to achieve the minimum standard of compliance which can be monitored through audits, fines for noncompliance or reports (Miller et al., 2020). Companies may voluntarily choose to exceed the obligated CS requirements and commit to further sustainable development that considers the interest of all stakeholders within and outside the business operations (Funk, 2003; Homburg et al., 2013; Raghubir et al., 2010; Wolff et al., 2020). This constitutes corporate social responsibility. At

the core of CSR lies the TBL framework. Hence, all CSR actions taken by the firm must reflect the three pillars of the TBL (Lock & Araujo, 2020). Typical for firms undertaking CSR actions is implementing sustainability management which is outlined by e.g., life-cycle assessments, sustainability reports, cleaner production measures, dialogue-based management and redesign of products with more sustainable alternatives (Schaltegger & Burritt, 2018). The United Nations Global Compact define six steps for companies for sustainable development strategies: commit, assess, define, implement, measure and communicate sustainability targets. (United Nations Global Compact et al., 2019).

2.1.3. Impact of Sustainability on Business Performance

Over the years there has been an increase in CSR activities and philanthropic engagement (Homburg et al., 2013; Schaltegger & Burritt, 2018). Around 90 percent of Fortune 500 companies have executives and departments solely dedicated to CSR (Homburg et al., 2013). Prior research has shown that sustainable actions have a beneficial long-term effect on the company value and secure a competitive advantage since it can decrease manufacturing costs, build a positive reputation which aids crisis management and attract talent (Du et al., 2010; Miller et al., 2020; Sanchez-Chaparro et al., 2022). Key drivers for this change are the pressure stakeholders placed on firms along with reputation management and the pursuit of shareholder value growth (Cone Communications, 2017; Cone Communications, & Ebiquity, 2015; Sanchez-Chaparro et al., 2022). Consumers expect companies to produce high quality products and services while simultaneously pursuing societal and environmental values. Moreover, they are demanding improvements in business practices and for companies to address social matters drive positive change (Miles & Covin, 2000). If companies can manage to conform to their expectations they can gain the consumer's trust which in turn can increase buyers intent, in addition to higher willingness to pay, and positively influence the overall attitude towards their business (Klein & Dawar, 2004; Sanchez-Chaparro et al., 2022). Should companies fail to even meet CS standards, consumers go as far as to boycotting the company altogether (Cone Communications, & Ebiquity, 2015; Porter & Kramer, 2006). Thus, stakeholder management is crucial for successful sustainable development. At times it can be challenging as integrating all stakeholder interests requires compromises between the groups that could potentially lead to greenwashing strategies to appease all parties involved (Engert et al., 2016; Sukitsch et al., 2015).

Whenever new policies concerning CS are introduced by the legislation, companies need to find new ways to incorporate them into their business strategy. More so, government agencies and strategic partners evaluate companies by their rating and brand value when considering alliances. Companies can improve their rating by engaging in CSR activities. Hence, it is crucial for companies to integrate sustainable actions into their operations (Kang et al., 2016; Miller et al., 2020).

In addition to consumers and legislation, shareholders play a significant role when it comes to CSR. Considering that CSR activities increase brand value, shareholders are strongly invested in advancing sustainable development as an increase in brand value leads to an increase in shareholder value. (Miller et al., 2020; Mishra & Modi, 2016).

Despite CSR possibly bringing in negative temporary cashflows, CSR shapes a company's brand positioning and corporate identity, legitimizes their existence claim and overall increases their profitability long-term (Mann et al., 2021; Shinkle & Spencer, 2012).

2.1.4. Sustainable Development in the Automotive Industry

Hitherto the automotive industry has had a considerably ambiguous relationship to sustainability (Shinkle & Spencer, 2012). This is owed to their significant impact on carbon emissions due to their value-adding activities, their customer's use of their products but also their possible positive influence on economic development (Cone Communications, 2017; Mayyas et al., 2012; Shinkle & Spencer, 2012). With global warming and resource depletion, the industry faces one of its greatest challenges. Therefore, they must modify their modus operandi to meet policy makers demands on reducing greenhouse gas emissions and environmental protection but more importantly conform to societal expectations (Sukitsch et al., 2015). This is imperative as the automotive industry is under the highest scrutiny for sustainable development as it is one of the highest CO2 producers (Wolff et al., 2020). In fact, a strategic reorientation can support their moral legitimacy and reputation. Additionally, willingly implementing CSR strategies can reduce threats of regulation imposed on the automotive industry (Russo-Spena et al., 2018).

It is important to mention that the automotive industry holds a major influence on sustainable development, having value chains that reach and connect multiple regions. If committed the industry could become a key driver in advancing sustainable development and reshape its reputation (Mayyas et al., 2012; Wolff et al., 2020).

2.2. Sustainability Marketing

2.2.1. Functions of Sustainability Marketing

As stakeholders' interest in CSR activities is growing and sustainability is becoming crucial for business survival, financial performance and overall a company's corporate identity (Du et al., 2010; Mann et al., 2021; Miles & Covin, 2000), transparency and credibility are key for CSR (Cone Communications, 2017). To achieve that, companies utilize sustainability marketing and public communication. Both tools lead to higher brand visibility, enhanced reputation, and trust which in turn builds credibility for CSR actions (Becker-Olsen et al., 2011; Mann et al., 2021; Miles & Covin, 2000; Raghubir et al., 2010). More so, the reputational advantages of authentic CSR and sustainability communication are vital to create long-term brand value (Mish & Scammon, 2010). When companies share their actions and points of view on social issues, they can gain the customer's support and even increase chances of customers switching to their brand and their willingness to pay (Mann et al., 2021; Sanchez-Chaparro et al., 2022). Likewise, sustainability communication is as equally important for shareholders and government regulators as it is for customers. Authentic CSR reporting leads to higher ratings which reduces perceived risks that can result in favorable discount rates and increases shareholder value. Additionally, it can minimize regulatory compliance activity and build long-term relationships with government regulators that allow them to contribute to developing new environmental regulations and receive concessions (Mann et al., 2021). Being able to get involved in new policies is highly relevant since government policies can shift consumers environmental behaviors (Kalamas et al., 2014; Minton et al., 2012).

As illustrated, sustainability marketing and communication are essential to building stakeholder trust and performance. Both must be authentic and truly reflect the company's image. Otherwise, it could raise skepticism among stakeholders and companies could run the risk of damaging their reputation.

There have been scandals such as the Volkswagen Dieselgate, where companies make use of deceptive communication (Kang et al., 2016; Siano et al., 2017). Siano et al. (2017) describe it as a new form of greenwashing. Consequently, it is when companies deliberately promote promising sustainable projects with no support and make false claims that they are not able to carry out (Porter & Kramer, 2006; Siano et al., 2017). This can be in form of selective or inaccurate disclosures and incomplete comparisons. Companies may resort to these strategies as it can be challenging to pursue all stakeholder interests simultaneously at times. Regardless, they must advert from these strategies as it could them at risk of severely damaging their reputation that they might not recover from. For example after the Dieselgate got public VW's stocks crashed one day by 22% (Li et al., 2018).

2.2.2. Channels of Sustainability Communication

To build stakeholder trust and achieve effective communication, companies need to directly engage with stakeholders. Financial disclosures and NFDs are integral tools for companies to report their actions and are considered as standardized frameworks for accountability (Kountouri et al., 2019). Prior research indicates that one of the most effective ways to communicate a company's CSR actions is through their sustainability reports which are usually independent from the company's annual reports (Lock & Seele, 2016). They provide additional information to stakeholders about the company's actions aside from their financial performance. If utilized responsibly, NFDs can facilitate building up a company's moral legitimacy and help gaining stakeholders' trust.

Aside from the traditional sustainability communication instruments, are companies using new media platforms like Twitter to engage with stakeholders and disclose CSR actions. Specifically, consumers are drawing information from social medial (SM) platforms. This is due to information being more accessible on SM platforms as well as less complex to comprehend (Lock & Araujo, 2020). Additionally, it gives consumers the opportunity to voice their opinions and directly address concerns to companies (S. Lee & Cho, 2011). From the company's perspective, do SM platforms enable dialogue and connecting with consumers on a personal level (K. Lee et al., 2013). In particular does the platform Twitter allow two things: connecting with consumers in form of bidirectional sharing and mass broadcasting. If companies succeed at this form of communication, it allows them to restore the consumer's trust after scandals and rebuild or improve their reputation.

To what degree SM platforms are effective to communicate with consumers depends on culture and consumer behavior (Minton et al., 2012). Minton et al. (2012) research reveals that SM use among consumers and their attitude towards sustainability varies between cultures. More so they suggest that sustainability marketing should acknowledge these differences in locations to successfully communicate sustainability to consumers.

3. Methodology

This research takes form of a latent semantic analysis (LSA) conducted with the program Python. LSA, a computer-aided topic modeling method, is a natural language processing technique that enables the identification of relationships between a set of documents as well as between terms within them (Kountouri et al., 2019). Namely, LSA highlights underlying conceptual meanings by uncovering similarities and differences in term and phrase usage. Thus, it produces textual themes in data (Deerwester et al., 1990). Hence, LSA overcomes problems of polysemy (a words having multiple meanings) and synonymy (multiple words sharing the same meaning).

By applying LSA, one can reduce the risk of bias related to subjective interpretation since no human coding concerning the topic themes is necessary. Hence, computer-aided identified themes are more reliable without having to predetermine categories of interest.

There has been broad use of quantitative content analysis ranging from automated essay grading to fields of medical or financial analyses, which shows how beneficial this application is for a vast array of fields of research. However, this has rarely been applied to the field of business ethics and sustainability (Lock & Seele, 2016). Kountouri et al. (2019) research of CSR is one of a few that proves that LSA can successfully be used as a tool for textual analysis.

The study focuses its analysis solely on the automotive industry. This industry was selected for multiple reasons. For one, external communication is becoming essential for companies in view of economic challenges like climate change mitigation and resource depletion. (Mayyas et al., 2012; Russo-Spena et al., 2018; Wolff et al., 2020). Secondly, their operations and strategic relations stretch around the entire globe. With their far-reaching value chains, they have the ability to influence sustainable development of entire regions.

For this study, particularly to facilitate the analysis, the assumption of designated communication channels for specific stakeholders has been made. Therefore, official reports, e.g., annual reports, are analyzed for shareholder and legislator communication and social media as a communication channel for consumers.

Additionally, this study includes a cross-country (UK and USA) component for the consumer analysis since consumer behavior and their sentiment towards sustainability depend on culture (Minton et al., 2012). Hence, it is of interest to examine if sustainability focuses vary within the consumer communication.

3.1. Data Collection

Data from the past three years was collected from four German automotive companies: Mercedes Benz, Audi, BMW, and Volkswagen. Since no complete reports of the year 2022 have been released yet, the time period was set from 2019 to 2021 . There are various reasons for the selection of the timeframe and sampled companies. Firstly, it is of interest to examine if CSR communication changed throughout the years. Secondly, the chosen companies have been selected due to their dominance in the automotive industry, especially in Germany. Data was only extracted from their official pages. These include annual reports, CSR reports, NFDs, and posts Twitter (see Table 1). In total, the sample consisted of 71 disclosures and 45302 Twitter posts. Moreover, two data sets, one from the United Kingdom (UK)and the other from the USA (US), were collected for each company's communication through Twitter. This cross-country examination was conducted as past research shows that cultural differences are reflected in consumer behavior (Becker-Olsen et al., 2011). For example, do consumers place different weighing on environmentally friendly goods. Hence, this research assumes that sustainability content might differ, depending on what region the company wants to reach.

3.2. Data Analysis

This section provides a description of the necessary steps of a LSA. Further, section 3.2.3 explains how the outputs were interpreted. All detailed Python code sheets as well as the outputs can be found in the appendix (see A1).

3.2.1. Preprocessing Data

The analysis was conducted using the software Python. Before applying LSA, the data was cleaned and prepared. First, all text from reports, annual reports, CSR reports and NFDs, was extracted and compiled into text (txt.) file format to analyze them in Python. The same process was done for social media posts. For the next steps the "nltk" package was applied, which is a natural language processing package that enables natural language processing analyses with Python, to clean text files (corpus) for further analysis. This cleaning process encompassed the following steps:

1. All text formatting was removed. This includes pronunciation, special characters, capitalization, and digits.
2. The text was tokenized into small units with "nltk.tokenize". Thus, the corpus was split into single terms which was important to classify the terms later on into topic themes.
3. From the tokenized data, all stop words were eliminated. Stop words are defined as terms that do not contribute to the underlying context of the corpus or do not hold any value. They function as neutral values. Additionally, all words that are extremely common were removed as well. Stop words typically entail names, articles, prepositions, days, weeks, months, and pronouns, e.g., she, hers, Friday. The complete stop words list can be found in the digital appendix (see A15).
4. Next, the text was stemmed, which is the process of finding the root of a word, with Porter's algorithm. The goal of stemming is to reduce derivationally related forms of a word to a base form. As an example, words cat, cats, cat's, and cats' would be returned to cat. Porter's algorithm is made up of five phases of word reduction. To visualize it, this would mean sses would become ss, ies to i, ss to ss and s would be eliminated. For example:
    - sses: caresses = caress
    - ies: ponies = poni
    - ss: caress = cares
    - s: cats = cat
5. In addition to stemming, lemmatization was applied to reduce inflectional related forms of a lemma. Lemmatization is a process of finding the form related of the related word in the dictionary. For example, with the tool "WordNetLemmatize" the token "wrote" would be returned to "write".

Table 1: Data sample

Company	Annual report	Sustainability report	Disclosures	Twitter (USA)	Twitter (UK)	Total
Audi	3	3	19	6482	3712	10219
Volkswagen Group	3	3	11	1071	6517	7605
Mercedes-Benz	3	3	15	5418	6673	12112
BMW	3	3	6	4718	6213	10942
Total	12	12	51	17689	23115	40879

3.2.2. The LSA Algorithm

The algorithm starts with employing the preprocessed data (d = number of compile files, with i = report) as the $[t_i \times 1]$ term-frequency vector. Each element of the vector is equal to how often the unique term was used in the document. The elements are calculated as the number of occurrences of the term divided by the total number of terms of the document. This term-frequency vector is applied since it is understood that if a word is more frequently used in a text, it should contribute more to the content and meaning of given text.

The term-frequency vectors are linked together afterwards. Therefore, common terms were only maintained once they formed the following matrix:

$$t \times d \tag{1}$$

where $t = \max\{t_i\}$, $i = 1, \ldots, d$ is the number of unique terms across the corpus.

Consequently, the rows of the matrix represent each unique term in the corpus and the columns represent each data sample. The frequency rates of every unique term make up the elements of the matrix. The term-document matrix (X) was built by reducing the elements by subtracting the corresponding row mean.

The process models each data sample in vector-space. Hence, each data set was decomposed into a simple vector of unique term frequencies. By applying singular value decomposition, LSA sets itself apart from regular word counting and ranking. Singular value composition reduces the dimensions of the matrix. This is also called noise reduction. Since polysemy and synonymy basically provide the same underlying information, more unique terms will be present. Singular value composition eliminates that noise and henceforth extracts the main underlying textual themes within a text. The themes are extracted from the covariances between the term frequencies of a text. Singular value composition therefore breaks the term-document value down into

$$X_{t \times d} = U_{t \times d} \times S_{d \times d} \times V^t_{d \times d} \tag{2}$$

$\boldsymbol{U}$ is the contribution of each unique word to the themes. The $\boldsymbol{V}$ columns are the principal values of $\boldsymbol{X}$, which are defined as the contribution of each theme to each document. Lastly, $\boldsymbol{S}$ describes the matrix of singular values. With this matrix, it is possible to extrapolate the importance of each theme to the text corpus in decreasing order. Figure 1 visualizes the above-mentioned function of singular value decomposition.

For the LSA algorithm the "genism" package available for Python was used. This package enabled to model LSA.

3.2.3. Interpretation and Evaluation of Themes

After retrieving the topics from LSA, the topics were evaluated and interpreted by the terms connected to each topic. For this interpretation it is important to mention that the positive values for each term do not necessarily mean that they have a positive or negative attachment to the topic. The values are determined by the vector of the SVD matrix. Since vectors can be both negative or positive, this means that the values of the terms are either positive or negative depending on the vector. The importance lies in considering the absolute value of each term. The higher the value the stronger the importance of the term to the underlying meaning of the text context.

After retrieving the theme topics or main trends of the LSA model, the topics were interpreted and evaluated by classifying them into the three categories: economic, environmental, and social. These categories are rooted in the triple bottom line framework by (Elkington, 1998). The topic classification was executed for each of the three stakeholders of the respective companies. As discussed in Section 2.1.1, the core of the TBL lies in striking a balance between the three pillars to achieve sustainable development. Rightfully so, the companies were assessed by how balanced the topics for each of their stakeholder communication channels were distributed across the three categories. Therefore, more evenly distributed topics for each channel, imply a higher degree of successful sustainability communication.

It is important to note that negative or positive coefficients in the LSA decomposition are not necessarily meaning negative or positive attachment to a topic since the basis can also have negative vectors. What is true is that a higher absolute value usually means a stronger relationship with the corresponding vector. Therefore, the term adds more value to the context of the content.

4. Results

4.1. Shareholders

This section will highlight the most important findings of the LSA analysis. As mentioned, each theme was classified

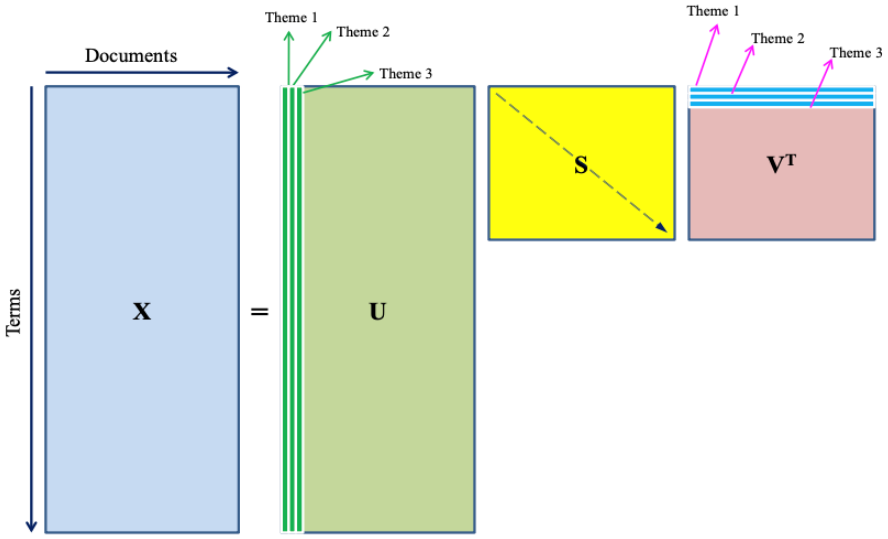


Figure 1: Singular value decomposition (retrieved from Kountouri et al. (2019))

into the three categories "economic", "environmental" and "social" by evaluation and interpreting the 50 terms and their values that comprise each topic. The following sections will highlight the most important findings for each stakeholder for each channel of the respective companies and summarize the topics in a comprehensive manner. For further insights please refer to section A3-A14 of the appendix which includes tables with all topic themes including five terms with the highest value. The list with all terms that were used for interpretation is available in the digital appendix (see A15).

The sample for shareholders consisted of the companies' annual reports from the years 2019 to 2021. The average annual report comprised about 320 pages.

As one can see, Audi has the most fluctuation in length, this is due to the fact that Audi changed its format from 2020 to 2021 from vertical to horizontal which allows more space. Furthermore, it is to mention that as of 2020 Audi combined their annual report with their sustainable report. Generally, Table 2 shows that all companies release annual reports of similar length with BMW scoring the highest page average of 368 and Audi the lowest with 274.

4.1.1. Mercedes-Benz

The coherence value (CV), which indicates how many topics are needed to explain the underlaying context of a text (Table 3), determined 18 topic themes. From the 18 topics, 16 were identified under the economic aspect with two falling under the environmental classification. Therefore, about 89% of the annual reports of Mercedes-Benz report on economic aspects and economic performance. The majority of environmental topics outline the themes financial performance compliance and reporting, and market value. Additionally, from Table 3 one can identify that the most frequently listed terms were "finance". "statement" "manage", "busi" (meaning business, which was shortened due to the stemming tool) and "market" for environment but also in total. To understand, how the topics were interpretated into themes and then classified into topics the following demonstrates an exemplary interpretation using the 10^{th} highest terms of topic 1.

Example:

> Topic 1: 0.768*"finance" 0.417*"statement", 0.414*"consolid" 0.156*"note" 0.075*"manag" 0.071*"risk" 0.066*"asset" 0.059*"servic" 0.057*"instrument" 0.047*"liabil" 0.043*"posit"

All terms are components of a balance sheet that are used to evaluate a company's financial performance. Hence, the theme for topic 1 would be labelled under "financial performance" which then will be classified under the economic aspect of the TBL since the financial performance focuses on e.g., investment returns, liquidity and revenue.

When addressing environmental aspects, Mercedes primary objective lies in reducing carbon emissions and supply chain changes and/or management. Interesting to mention is that these terms are listed in combination with compliance. Thus, it is difficult to determine if these fall under CR or CSR actions. More so, no topic with social aspects could be identified

4.1.2. Audi

Through the LSA analysis three topics were identified. The determined topics were classified into the economic and social categories, with two belonging to the former. One of them comprises financial performance with terms including "asset", "liabil", "statement", "instrument" and "risk". The other was identified as business reorientation with terms like "risk", "manage", "opportunity", "market" and "product". This can be interpreted as development in new products that can lead to capturing higher market value. Lastly, the topic concerning social aspects includes "vote", "right", "law", "exceed"

Table 2: Annual reports

Company	2019	2020	2021	Average
Mercedes-Benz	350	267	348	322
Audi	274	381	138	264
Volkswagen	354	342	398	365
BMW	368	262	353	398

Table 3: Mercedes-Benz shareholders: topic themes

	Economic	Environmental
Terms with high frequency	"finance", "statement", "manage", "busi", "market"	"carbon", "emiss", "market", "vehicle", sustain"
Themes	Financial performance, reporting, compliance	Carbon emission, change in supply chain
Number of topics	16 (88,9%)	2 (11,1%)

and "human". This was understood as integrating different stakeholders and exceeding laws and policies which include human rights. Since the social aspects are defined as actions contributing to the community, this topic was classified as social. For better visualization Table 4 summarizes these findings, indicating the identified themes, the highest value terms and number of topics with overall percentage.

4.1.3. Volkswagen

Results for Volkswagen show a CV of 13 and correspondingly 13 topic themes. All 13 themes were identified as part of the economic class. Like the companies above, financial performance was one of the most often identified themes, in total 7 out of 13 times. Frequent terms being "finaci", "hedg", "risk", "asset","statement", "tax" and "market".

Further the themes sales, market and brand value, and product/ car development were identified (see Table 5). They were both identified by two topics. Terms for sales include "sale", "market", "commerci", and "vehicl". The terms"Brand", "market", "credit" were interpreted for brand and market value. The two topics outlining car development contained the terms "car","vehicle", "develop", "passeng", and "chang".

4.1.4. BMW

For BMW, the CV indicated a topic number of 24. 17 were classified under the economic aspect and 11 under environmental with none being identified as social. Four of these topics (topic 3, 4, 15 and 21) overlapping. The overlapping topics mainly outline how environmental measures such as reduction in carbon emission in car development or sustainable resource (topic 3, 15) use, impact BMW's financial performance. Again, results show that financial performance is the most frequent theme, with four topics (topic 1, 7, 9, 10) being identified as such. Most frequent themes for the environmental category are sustainable supply chain (2 topics) and development of sustainable cars/ products. Interestingly, more diverse themes were found, ranging from carbon emission footprint of production (topic 11) to alliances and joint ventures (topic 13). This is because a wider variety of terms was detected by the LSA. Table 6 shows the most frequent terms that were determined through the LSA.

4.1.5. Cross-company comparison

As results showed, none of the companies' annual reports reflect all three aspects of the triple bottom line. Volkswagen showed the lowest performance by only representing the economic aspect in their annual report (see Table 7). The only company that had at least one theme classified as social is Audi. On the other hand, Mercedes-Benz and BMW show themes in the environmental category. Although, it is important to mention that BMW demonstrate more balanced results with 45,8% of its overall topics being classified as environmental compared to Mercedes' 11,1%.

4.2. Legislation

In this section, results for sustainability reports and other NFDs are shown which are considered for the analysis of the communication between legislation and the respective companies. Overall, the average sustainability report is about 159 pages long. As seen in Table 7, from all companies Mercedes has the lengthiest reports with an average of 232 and BMW the shortest with about 87 pages. Notably, Volkswagen has been the most consistent in report length averaging with 103 pages.

4.2.1. Mercedes-Benz

Three topics were defined by the CV. Two of them (topic 1 and 3) were classified as environmental aspects of the TBL (see Table 8). The first one contains terms such as "sustain", "manage", "protect", "human", "emiss" and "environment" and thus, was categorized as sustainability management and environmental action within the business operation. Topic 3 includes terms like "manage", "emiss", "product", "sustain", "material and "vehicle" which was outlined as the theme development of environmental-friendly products. Lastly, Topic 2 was classified under the aspect social outlining the theme employee protection and work environment. Terms included are "sustain", "manage", "employee", "human", "train" and "protection".

Table 4: Audi shareholders: topic themes

	Economic	Social
Highest value terms	"fianci", "statement", "product", "asset", "risk"	"vote", "right", "amount", "attribute", "law"
Identified theme	Financial performance,	Policies/stakeholder integration
Number of topics	2(66, 7%)	1(33, 3%)

Table 5: Volkswagen shareholders: topic themes

Financial performance (topic 1-3, 5, 8, 10, 12)	"financi", "asset", "statement", "hedge", "consolid",
Brand and market value (topic 5, 11)	"brand", "market", "vehicle",
Sales (topic 4, 9)	"sale", "manage", "revenue", "market"
Car / product development (topic 7, 13)	"car", "vehicle", "develop", "passeng",

Table 6: BMW shareholders: topic themes

	Economic	Environmental
Terms with high frequency	"finance", "statement", "asset", "busi", "market", "risk"	"carbon", "emiss", "market", "service"," vehicle", sustain", "develop"
Most frequent themes	Financial performance, reporting, compliance	Carbon emission reduction/ compliance, change in supply chain, product development
Number of topics	17 (54,2% single; 70,8% incl. overlapping themes)	11 (29,2% single, 45,8% incl. overlapping themes)

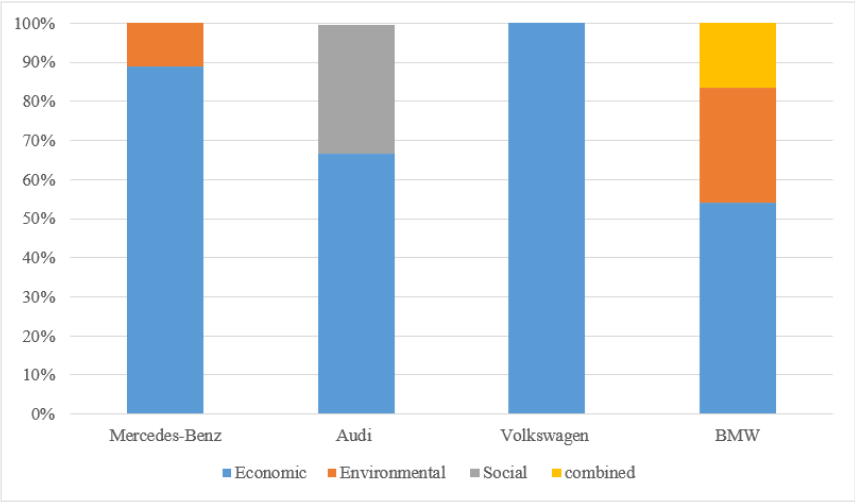


Figure 2: Cross-company comparison of shareholder communication

Table 7: Sustainability reports

Company	2019	2020	2021	Average
Mercedes-Benz	205	193	298	232
Audi	127	381	139	215
Volkswagen	100	97	111	103
BMW	142	74	45	87

4.2.2. Audi

With a CV of 13, four topics (topic 2-4, 8) were classified as social, another four as environmental (topic 1, 5-7) and five (topic 9-13) as economic. The environmental themes mainly outline reduction of carbon emission of business operation and products, while social themes include work culture and employee right, and forming partnerships with various partners (see Table 9) Lastly, the five topics classed as

Table 8: Mercedes legislation: topic themes

	Environmental	Social
Terms with high frequency	“sustain”, “manage”, “protect”, “emiss”, “employe”	“sustain”, “manage”,” employee”, “compliance”, “risk”
Themes	Sustainability management, sustainable product development	Employee protection, work environment
Number of topics	2 (66,7%)	1 (33,3%)

economic comprise terms like "vehicle", "strategi", "profit", "busi", "oper" and "suppli". Thus, leading to the interpretation of the themes production and manufacturing, profitability and business strategy. Overall results show a fairly even representation of all aspects of the TBL in Audi’s sustainability communication to legislation actors.

4.2.3. Volkswagen

Results show that 9 topics were identified by the CV of the LSA. From these, 7 topics were interpreted as aspects of the environmental class. One of them, Topic 5, was categorized under both classes, social and environmental, and outlines the theme environmental action with integration of the public and partners. Terms include are "human", "busi", "emiss", "integr", and "partner". Main themes that fall under the environmental class are sustainable business/ management, the implementation of sustainable supply chain management and environment/ climate-conscious car production. Frequent terms of this class include "sustain", "chain", "suppli, "emiss", "human" and "busi".

Furthermore, the other theme that was identified for the social category describes the communication and integration of the public in the company’s actions and performance. Terms linked to this theme for instance are "inform", "public", "directly"," complianc", "integr" and "partner" which indicate that companies disclose their compliance to the public and consider various stakeholder groups into their business operations.

Lastly, the only theme that was able to be interpreted as economic is topic 7 which outlines the impact of green finance and sustainability on the financial performance. This interpretation was based on the LSA assigning high values to terms like "financ", "green", "bond" and "portfolio".

4.2.4. BMW

For this group, 17 topics were able to be identified. From these 14 (topic 1, 2, 4, 6-10, 12-17) have been classified under the environmental aspect of the TBL. Main themes outline the reduction of carbon emission and the use of renewable energies as well as climate conservation action. Terms often included in these topics are "emiss", "carbon", reduc", and climate" (see Table 10) Notably, environmental themes make up about 80% of the entire data set.

Further topic 13 and 15 have been classified as both economic and environmental, were interpreted as conforming to GRI regulations and reaching climate targets. Reasoning behind the combined classification is, that from those term it was not possible to determine whether these are CR or CSR actions since deciphering the motivation behind the environmental action was not feasible. Hence, both topics were given both classifications.

Moreover, two topics of the social classification were identified (topic 3 and 5) which were interpreted as themes about employee care and the integration and exchange between the company and the public. Notable terms belonging to this group are "societi", "employe", "sustain", and "manage". Lastly, another economic theme, besides the topic 3 and 5, was "identified". Topic 11 outlines the theme product procurement and supply chain. Terms linked to it are "target", "supplier", "purchase", "scope" and "suppli".

4.2.5. Cross-company comparison

Overall, results show that Audi is able to achieve a balance in addressing all three aspects of the triple bottom line, with a near equal distribution of the TBL classes (seen in Figure 3). Important to highlight is that all companies, except for Mercedes-Benz, address all three aspects to some degree. From Figure 3 one can see that Volkswagen and BMW show similar classification distributions. Although the combined classification of VW refers to the social and environmental category. Whereas for BMW, it addresses the combination of environmental and economic aspects.

Notably, for all four companies the environmental aspect dominated in the legislation communication. Most frequent theme addressed is the reduction of carbon emission of production and vehicles, and climate-conscious production and supply chain management, and the use of renewable energy. For the social classification, employee care and work environment are most frequently addressed as well as the integration and dialogue with the public and various stakeholders.

Lastly, for the economic side, companies most frequently address the financial performance and the impact of environmental activities on it.

4.3. Consumers

4.3.1. Mercedes-Benz

As can be seen in Table 12, the UK channel is more active than the US channel with about 500 more Twitter posts. In total Mercedes-Benz’ UK. Twitter account released 6673 posts within the three years whereas the US account posted a total of 5498 posts. Moreover, from the LSA of US consumers communication the derived coherence value indicated a topic number of 3. Two of them were interpreted as economic. Majority of terms included are names of car models e.g., "gle",

Table 9: Audi legislation: topic themes

	Environmental	Social	Economic
Terms with high frequency	“sustain”; “manage”, “employee”, “emiss”, “develop”	“employee”, “sustain”, “society”, “emiss”, “manage”	“vehicle”, “strategi”. “oper”, “compliance”, “suppli”
Themes	Carbon reduction, sustainable business operation	Work culture, dialogue with partners, partnerships	Production/manufacturing, business profitability
Number of topics	4 (30,8%)	4 (30,8%)	5 (38,4%)

Table 10: Volkswagen legislation: topic themes

	Environment	Social	Economic
Terms with high frequency	“sustain”, “emiss”, “integr”, “suppli”, “busi”	“employee”, “sustain”, “human”, “suppli”, “partner”	“green”, “finance”, “framework”, “bond”, ”elig”,
Themes	Sustainable business/ supply chain, climate-conscious prodcution, resource/ raw material	Public and stakeholder integration reporting,	Financial performance with sustainable actions
Number of topics	7 (66,7% single, 77,8 % combines)	2 (11,1% single, 22,2% combines)	1 (11,1%)

Table 11: BMW legislation: topic themes

	Environment	Social	Economic
Terms with high frequency	“emiss”, “carbon”, “reduc”, “climate”, “suppli”	“employe”, “societi”, “manage”, “sustain”, “train”	“suppli”, “chain”, “risk”, “supplier”, “gri”
Themes	CO2-emission reduction, renewable energy, sustainable supply chain	Employee conditions, integration & interaction with stakeholders	Meeting targets in business operations and supply chain
Number of topics	14 (70,6%), 16 (82,3 % combined)	2 (11, 8%)	1 (5,9%), 3 (17,6% combined)

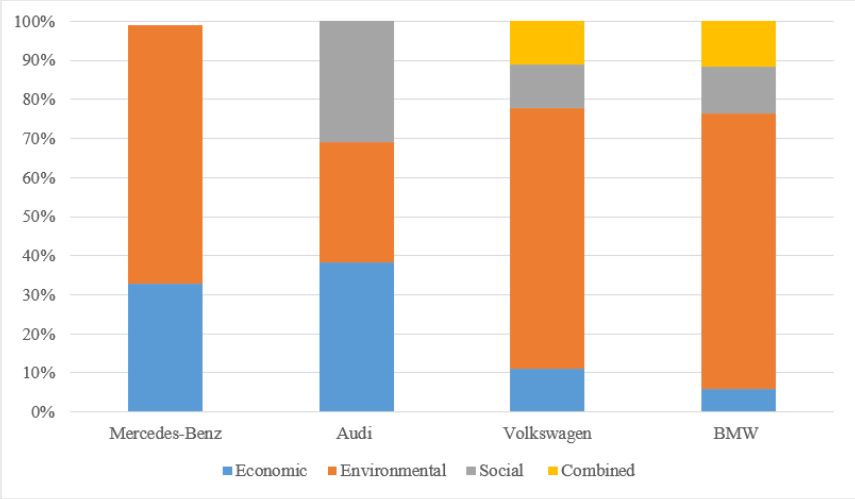


Figure 3: Cross-company comparison of legislation communication

"gclass", and components of cars. Thus, this is construed as advertisement and promotion of products. Topic 2 describes customer service with terms such as "team", "custom", and "assist". Therefore, it was classified under the social aspect of the TBL.

For the UK consumer channel, a coherence value, and hence a topic number of 6 was identified. All six topics fall under the social aspect and outline the theme customer service and public audience engagement. Most frequent terms include "team", custom", "service", "reach", "thank", and "email", indicating communication between customers and the company is primarily about customer inquiries.

Table 12: Mercedes- Benz consumers: topic themes

	US consumers (n=5498)		UK consumers (n=6673)
	Economic	Social	Social
Terms with high frequency	“gle”, “sclass”, “allnew”, “design”, “mbambassador”	“team”, “custom”, “hear”, “sorri”, “assist”	“team”, “service”, “custom”, “contact”, “please”
Themes	New product advertising, Product innovation	Customer service/ Communication with community	Customer service/ Communication with community
Number of topics	2 (66,6%)	1 (33,3%)	6 (100%)

4.3.2. Audi

Audi’s results for its customer communication show that Audi’s UK Twitter account is almost twice as active as its US counterpart. In total Audi UK posted 6483 tweets and Audi US 3712 tweets. For the US channel a CV of 2 was determined with both topics outlining customer service and strengthening the relationship to the public. Therefore, the topics were classed as social. Terms that led to this interpretation are "pleas", "inform", "reach", "thank", "assist" and "conveni".

Similarly, all identified topic themes for the UK analysis were classified under the social category. In total the CV indicated 5 topics. All topics include similar terms that also outline the theme customer service and strengthening public relations. As seen in Table 13 topics include terms such as "pleas", "hear", "thank", "assist", and "concern".

4.3.3. Volkswagen

The total number of tweets posted on the US account are 1071 posts which is comparatively a small fraction of the UK account with 6517 posts. Nevertheless, results show more variety within the topics of the US sample. With 18 topics in total, 13 (topic 1, 2, 5-12, 14, 16, 17) of them, making it almost three quarters of the total, were classified into the economic category. Most common themes of the 13 topics are product and brand promotion as well as announcements of new products. These themes were identified by the high occurrence of car model names such as "jetta", "tiguan" and "atla". Besides car models, terms included are "sale", "latest", "announce", "concept" and "celebrat".

Additionally, three (topic 3, 4, 18) environmental and two (topic 13, 15) social themes have been identified. Environmental themes addressed product changes with environmental impact. Terms that were included are "mobil", "future", "electric", "sustain" and "batteri". Both social themes illustrate forming partnerships and integrating various actors into the research & development process as well as educational programs. Terms included and used for interpretation are "join", "hub" "research", "student" and "learn".

In contrast, all 16 topics for the UK channel fall under the social category with the main theme being customer service and communication to the public and community, with terms like "please", "thank", "team", "touch ", and "contact".

4.3.4. BMW

23 topics for the data set for US consumer communication, 2 themes of the economic category were identified. Terms of the two topics show that BMW tweet economic content for sales promotion and information about car models. Included terms in the topics are "drive", "power", "experi" (see Table 15) The 21 remaining topics are compiled under the theme customer service and community engagement.

In contrast, results for consumer communication for UK indicate that all 26 topics are related to customer service and community engagement. Frequently listed terms are "please", "sorri", "thank and "team".

4.3.5. Cross-company and cross-country comparison

Overall, results and data sample show that the companies’ UK Twitter accounts are more active in releasing higher quantities of posts. Moreover, it is to highlight that majority of themes fall under the social aspect of the TBL with the most frequent theme being customer service. Further, UK Twitter content differs from US content within each company and across companies (see Figure 4). For example, does BMW US address economic and social aspects and BMW UK exclusivley social aspects. Hence, content released in one region does not automatically correspond with the content of other regions.

As seen in Figure 4, the majority solely address the social aspect. Only Volkswagen US results show that all three aspects are addressed. Important to highlight is that the environmental aspect is strongly underrepresented with only Mercedes-Benz US and Volkswagen US addressing environmental aspects.

5. Discussion

The purpose of this study is to investigate how companies use different channels to communicate sustainability efforts to different stakeholders. Correspondingly, the LSA method was applied to examine the communication of four automotive manufacturers. The results show that the companies primarily focus on disclosing its financial performance when addressing the economic side of the TBL to shareholders. If environmental elements are mentioned, it is only in correlation to its added value to the company’s financial operations. This demonstrated tendency to meet performance targets and communicate them to shareholders is coherent with what prior research (Miles & Covin, 2000; Schaltegger

Table 13: Audi consumers: topic themes

	US consumer (n= 3712) Social	UK consumers (n=6482) Social
Terms with high frequency	"please", "http", "best", "addit", "assist"	"pleas", "thank", "look", "good", "concern"
Themes	Customer service/ inquiries	Customer service/ inquiries
Number of topics	2 (100%)	5 (100%)

Table 14: Volkswagen consumers: topic themes

	US consumers (n = 1071)			UK consumers (n = 6517)
	Economic	Environmental	Social	Social
Terms of highest frequency	"sale", "suv", "product", "design", "announce"	"future", electr", "sustain", "drive", "develop"	"join", "design", "team", "hub", "learn"	"please", "team", "touch", "contact", "problem"
Themes	Product/ brand promotion, new product introduction	Product changes with environmental impact	Alliances and integration of different actors in research & development	Customer service, public communication
Number of topics	13 (72,2%)	3 (16,7%)	2 (11,1%)	16 (100%)

Table 15: BMW consumers: topic themes

	US consumers (n=4718)		UK consumers (n=6213)
	Economic	Social	Social
Terms with high frequency	"drive", "ultim", "coup", "power", "experi"	"please" "sorri", "team", "custom", "learn"	"please", "sorri", "hear", "thank", "team"
Themes	Sales promotion, Test drive	Customer service, community engagement	Customer service, public relations engagement
Number of topics	2 (8,7%)	21 (91,3%)	26 (100%)

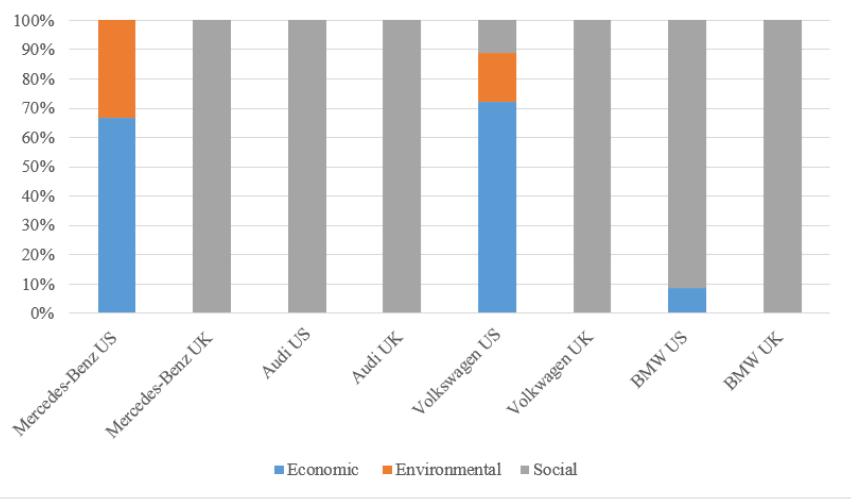


Figure 4: Consumer communication: Cross-country and cross-company comparison

& Burritt, 2018) has revealed. As annual reports are first and foremost catered to shareholders and investors, disclosing a company's financial performance shows to be the most relevant aspect for this group. This is due to the investors' and shareholders' motivation to maximize their own shareholder value. It can be observed from the analysis' results that terms such as "cost", "profit" and "cash" are linked to the economic aspects. Like Raghubir et al. (2010) and Russo-Spena et al. (2018) mention, can CSR actions lead to short-term negative cashflows but can also have a long-term positive effect

on the company's value. Companies must therefore proof to investors that the long-term benefits outweigh its short-term negative impact on cash flows. Through annual reports can companies justify their rationale behind taking CSR actions and highlight to shareholders that sustainable actions can secure the company's viability and with it its shareholder value.

Moreover, reports are largely being used as a tool in the decision-making of investors and rating agencies. Companies are motivated to primarily disclose its financial performance in regard to CSR as it has a direct influence on cost of capital. This is because investment decisions are partially based on what companies disclose (Clarkson et al., 2013; Richardson & Welker, 2001). Correspondingly, a possible reason for companies to focus on financial aspects is that measures and metrics are more coherent and standardized for financial performances (Elkington, 2018), which can decrease estimation risk. This facilitates rating agencies in their company evaluation. Because companies seek to improve their rating to attract investors and overall increase their reputation and brand value, setting their focus on financial performance shows to be the most effective strategy to achieve these targets.

Interestingly, only Audi touches upon the social aspect by outlining the importance of integrating various stakeholders such as legislation into their business decision-making and operations. It is possible that companies barely address social or environmental actions in isolation to shareholders because they might be more interested in what implications CSR action have on them instead of the motivation behind it. Hence, the financial outcome of these actions is emphasized in annual reports in form of cost of capital, profitability and most importantly its effect on shareholder value.

In contrast to the shareholders communication, results for the legislation communication indicate a more balanced communication. All companies, except for Mercedes-Benz, who merely discloses environmental and economic aspects, address all three aspects of the TBL.

Notably, the most covered category is the environmental category. Although it is important to highlight that for this analysis NFDs were used. Hence, its content and targets are specifically designed to address sustainability and therefore all three pillars of the TBL.

The most common themes identified for the environmental aspect are efforts of reducing carbon emissions in production as well as car design, the implementations for more sustainable, climate-conscious supply chains, and the use of renewable energy. These results align with prior research that points out that the automotive industry faces pressure from stakeholders and stands under intense scrutiny due to being one of the highest carbon emission producers.

Hence, disclosing efforts in regard to carbon emissions reduction and climate preservation becomes vital to them to withstand pressure and conform to expectations from stakeholders. With policies and regulations in place for carbon emissions mitigation, companies address these issues in their NFDs to show compliance and possibly decrease the threat of new regulations or further compliance costs (Dhaliwal et al., 2011; Russo-Spena et al., 2018).

Additionally, showing compliance and more so communicating actions to exceed them, by taking steps towards sustainable development throughout the entire supply chain, builds trust and credibility. As Mann et al. (2021) points out, building credibility and especially gaining policy makers' trust offers companies the opportunity to build relationships with government regulators which possibly allows them to influence future decision-making regarding environmental regulations.

Similar to shareholder communication, it was observed that companies address their financial performance in connection to sustainable action to legislation. Other economic aspects they address are meeting operational targets. To gain trust and build credibility, it is essential to report actions truthfully and authentically (Mish & Scammon, 2010). Therefore, one can infer from the study's results that companies disclose their financial performance and benefits from CSR actions as they are in its core a business and hence, strive to stay profitable. Solely reporting CSR actions could raise skepticism from stakeholders and could be perceived as inauthentic.

More so, indicating that their business operations can generate positive outcomes on various levels, for example financially and environmentally, can increase the company's chances of forming alliances with government agencies (Kang et al., 2016).

The study's result also show that all companies address social aspects except for Mercedes-Benz. Though, it must be noted that the social aspect is overall strongly underrepresented with an average of just 16%. The only company that has been able to strike a balance and address all aspects in an equal manner is Audi. The dominant themes that the companies address are labor practices and fair operating practices. This shows that companies emphasize on implementing practices to improve their employees' work conditions and welfare. Terms such as "health", "employee", and "train" indicate that the respective companies introduce training programs for employees and health care policies. This aligns with Elkington (1998) classification of the social bottom line, which is defined as efforts towards better welfare and fair treatment of the community and stakeholders.

This study suggests that disclosing labor conditions and complying to labor policies has a similar effect on legislation as the compliance to environmental and CR regulations. If companies showcase efforts towards sustainable and ethical labor conditions, they can reduce the threat of new and stricter regulations being imposed on them. Further, this could attract new talent and improve their reputations as employers. Hence, CSR reports can additionally serve as recruitment tool by highlighting employee benefits to possible job seekers.

Equally important to stress is that the length of reports as well as the quantity of disclosures published does not automatically correlate with greater efforts towards sustainable development. This could be observed in this study. As an example, Mercedes-Benz published the most NFDs, yet it was

the only company that focused exclusively on environmental and economic concerns.

According to the consumer communication analysis, Twitter serves as an instrument to interact and connect with consumers. In contrast to other channels, companies are able to interact directly with the targeted stakeholder group. Notably, most interactions encompass customer inquiries about issues and customer service-related matters. Additionally, results show that the communication is bidirectional with oftentimes consumers initiating the conversation by inquiring about an issue that occurred to them. As a result of connecting with customers and responding to their questions, the company is able to build relationships. The study shows that Twitter can also serve as a customer service touchpoint and a forum for Q&A besides sharing information. From a consumer's perspective, social media, in this study Twitter, allows them to become proactive and initiate communication in contrast to the traditional communication where interaction is usually initiated by the companies in form of disclosure releases or posts. It is true that strengthening relationships with consumers and the community falls under the social bottom line, but a crucial part of the social bottom line is largely overlooked. Contributions to the community and social issues like labor conditions are not addressed. A possible rationale behind this could be that most communication that outline the social bottom line is initiated by consumers. Thus, they primarily revolve around customer service matters.

On the other hand, the results show that for the US economic content has been released that primarily focuses on promoting new car releases or innovations to current models. This indicates that Twitter is instrumentalized as an additional channel for advertising and marketing.

There is a noticeable neglect of environmental aspects of the TBL. Findings show that environmental aspects are only addressed in connection to new product released. However, the values of the term, e.g., "emiss", "carbon", relating to the environmental bottom line indicate a weak relation between the terms and the overall context of the topic. The use of Twitter by companies may be primarily for social interactions and for strengthening relationships with consumers. Therefore, it is possible that companies use other platforms such as Facebook or Instagram for more informational content since Twitter posts are limited to a maximum of 280 characters (K. Lee et al., 2013). This makes it difficult to appropriately address important issues that may require more context and space than 280 characters allow. As pointed out by S. Lee and Cho (2011), Twitter can be used to quickly resolve issues which in turn can be useful for crisis management. This is seen in this study as companies directly respond to customer inquiries and can therefore maintain customer satisfaction.

The cross-country analysis implies that the use of SM as well as the degree of which it is utilized depends on the region the company wants to reach. Results show that UK consumers are more active and outspoken about issues that occur. This is seen by the number of Twitter posts collected for the analysis. The number of tweets from the UK accounts exceed those of the US account. In addition, it can be inferred that UK consumers place a higher value on customer service. Meanwhile in the US emphasize more on informational content in regard to products and car models. Therefore, companies evaluate the region they operate in and tailor their communication to different audiences. This makes it interesting for future research to examine to what degree SM presence and its focus differ from region to region.

Coming back to Elkington (1998) triple bottom line framework a company is truly sustainable and committed to sustainable development if they incorporate all three aspects, economic, environmental, and social, into their business operations. After applying the framework to this study and examining each stakeholder channel, the results show that companies rarely address all three aspects in one channel. Results show that companies are particularly focused on one aspect of the TBL, depending on which stakeholder they wanted to address. It was observed that companies utilize annual reports to disclose their financial performance to shareholders who are motivated to increase their own shareholder value. Meanwhile for legislators, they heavily reported on environmental aspects such as actions towards climate preservation and carbon emissions reduction. As for the communication to consumers, companies focuse on customer care and building relationships with.

In light of these findings, it can be said that companies selectively communicate parts of sustainability to specific stakeholder groups. Companies must act with caution as a high level of selective disclosure can be perceived as manipulative "cherry picking", and only disclosing what is beneficial to them. This form of greenwashing could in turn severely damage their reputation (Porter & Kramer, 2006; Siano et al., 2017).

For this study and its analysis, the assumption was made that companies use one channel for a specific stakeholder. With this consideration, the results indicate that only legislation communication was capable of fulfilling the TBL and, thus, succeeding at communicating sustainability. Nevertheless, it is important to stress again that for this stakeholder group CSR and NFDs were used which are specifically designed to address all aspects of the TBL.

When considering the companies' communication as total sum of all their channels it is plausible to recognize that they attempt to address and communicate the TBL in an equal manner to stakeholders. Arguably, it is reasonable to suspect that companies use multiple channels to disclose CSR action to a stakeholder group which as was not investigated in this study and should be explored in future research.

5.1. Limitations and Future Research

As this study exclusively focused on textual content, findings cannot be generalized or applied in a larger context to their entire communication. Results in a different study could differ if for example images or even special characters are considered which are disregarded in an LSA. Therefore, future research should examine how textual content in connec-

tion with image content is used in sustainability communication.

Moreover, this research is restricted by computer-processor capacities. Data collection and sampling was therefore limited. A larger sample might yield more nuances and context to the results.

Nevertheless, this study shows that computer-aided quantitative analysis can be a powerful tool for textual context analysis. Notably, the topics identified by the LSA were afterwards human-encoded and classified into the three aspects of the TBL which exposed the results to the risk of bias as the interpretation of each theme is subjective. Nonetheless, this study shows that this risk is strongly reduced beforehand since no preliminary human-encoding is needed which is an essential step in traditional textual analyses.

Although LSA proves its benefits in this study, LSA is limited in capturing multiple meanings of a word. This problem could be overcome by additionally applying a latent dirichlet allocation method as it enables the detection of multiple meanings of a word. This would offer new research directions to widen the use of computer-aided textual analyses in the field of business ethics.

Because the study solely focuses on the automotive industry, results might be homogenous and could vary when applied to other industries. This should be further explored in future research.

6. Conclusion

The aim of this paper is to examine how companies communicate sustainability to different stakeholders. Under the assumption that companies use one channel for each stakeholder, a latent semantic analysis has been conducted to examine the communication of four companies of the automotive industry.

The study offers new insights into company communication and draws attention to the fact that companies are still not disclosing all aspects in one communication channel through their textual content. More so, it shows that companies use specific channels to disclose different aspects of their company and tailor them to the stakeholder they want to reach. Findings show that companies emphasize different aspects of the TBL when communicating to stakeholders. For shareholders, its focus is on the company's financial performance and its implications for shareholder value, therefore, addressing primarily the economic bottom line. On the other hand, environmental issues such as carbon emission reduction stand at the forefront of legislation communication. Lastly, the focus of the consumer communication lies on the social side of the TBL. In this case, companies prioritize on connecting with customers and its community through dialogue and interactions on social media platforms, here Twitter.

Drawing from the research findings, companies still lack the capability to address all aspects of the TBL in each channel. It is crucial for companies to find a way to address all three aspects of the TBL in all their communication channels individually and turn away from selectively reporting information to specific stakeholders. Otherwise, they run the risk of severe reputational damage.

This paper expands and contributes to the limited research (Kountouri et al., 2019; Liao et al., 2018; Lock & Seele, 2015) which applies quantitative methods for content analyses and offers researchers in fields of ethical business and sustainability new method approaches in addition to traditional qualitative methods. Moreover, it illustrates that computer-aided content analyses can reduce the risk of bias. Although further research needs to be conducted in terms of finding alternatives or extensions to LSA since this method is still prone to bias during interpretation of the topic outputs.

Lastly, this research highlights the strength of computer-aided methods that enables large data collection and analysis in a shorter timeframe. With a strong computer processor, a substantial amount of data can be collected. This would enable a cross-industry and multi-channel analysis with far reaching implications that can be applied to a wide range of industries and could facilitate companies to successfully communicate sustainability.

References

Beal, D., Eccles, R., Hansell, G., Lesser, R., Unnikrishnan, S., Woods, W., & Young, D. (2017). *Total Societal Impact: A New Lens for Strategy*. BCG Global. https://www.bcg.com/publications/2017/total-societal-impact-new-lens-strategy

Becker-Olsen, K. L., Taylor, C. R., Hill, R. P., & Yalcinkaya, G. (2011). A Cross-Cultural Examination of Corporate Social Responsibility Marketing Communications in Mexico and the United States: Strategies for Global Brands. *Journal of International Marketing*, *19*(2), 30–44. http://www.jstor.org/stable/41304393

Clarkson, P. M., Fang, X., Li, Y., & Richardson, G. (2013). The relevance of environmental disclosures: Are such disclosures incrementally informative? *Journal of Accounting and Public Policy*, *32*(5), 410–431.

Cone Communications. (2017). 2017 Cone Communications CSR Study. www.conecomm.com/2017-CSR-Study

Cone Communications, & Ebiquity. (2015). *2015 Global CSR Study*. Cone Communications. https://static1.squarespace.com/static/56b4a7472b8dde3df5b7013f/t/574efd15b654f9eda8bae855/1464794409693/2015_Cone_Ebiquity_Global_CSR_Report.pdf

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*(6), 391–407. https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9

Dhaliwal, D. S., Li, O. Z., Tsang, A., & Yang, Y. G. (2011). Voluntary nonfinancial disclosure and the cost of equity capital: The initiation of corporate social responsibility reporting. *The Accounting Review*, *86*(1), 59–100.

Du, S., Bhattacharya, C. B., & Sen, S. (2010). Maximizing Business Returns to Corporate Social Responsibility (CSR): The Role of CSR Communication. *International Journal of Management Reviews*, *12*(1), 8–19. https://doi.org/10.1111/j.1468-2370.2009.00276.x

Elkington, J. (1998). Partnerships from cannibals with forks: The triple bottom line of 21st-century business. *Environmental Quality Management*, *8*(1), 37–51. https://doi.org/10.1002/tqem.3310080106

Elkington, J. (2018). 25 Years Ago I Coined the Phrase "Triple Bottom Line." Here's Why It's Time to Rethink It. *Harvard Business Review*. https://hbr.org/2018/06/25-years-ago-i-coined-the-phrase-triple-bottom-line-heres-why-im-giving-up-on-it

Engert, S., Rauter, R., & Baumgartner, R. J. (2016). Exploring the integration of corporate sustainability into strategic management: a literature review. *Journal of Cleaner Production*, *112*, 2833–2850. https://doi.org/10.1016/j.jclepro.2015.08.031

Funk, K. (2003). Sustainability and performance. *MIT Sloan Management Review*, *44*(2), 65.

Goh, C. S., Chong, H.-Y., Jack, L., & Faris, A. F. M. (2020). Revisiting triple bottom line within the context of sustainable construction: A systematic review. *Journal of Cleaner Production*, *252*, 119884. https://doi.org/10.1016/j.jclepro.2019.119884

Homburg, C., Stierl, M., & Bornemann, T. (2013). Corporate Social Responsibility in Business-to-Business Markets: How Organizational Customers Account for Supplier Corporate Social Responsibility Engagement. *Journal of Marketing*, *77*(6), 54–72.

Kalamas, M., Cleveland, M., & Laroche, M. (2014). Pro-environmental behaviors for thee but not for me: Green giants, green Gods, and external environmental locus of control. *Journal of Business Research*, *67*(2), 12–22. https://doi.org/10.1016/j.jbusres.2013.03.007

Kang, C., Germann, F., & Grewal, R. (2016). Washing Away Your Sins? Corporate Social Responsibility, Corporate Social Irresponsibility, and Firm Performance. *Journal of Marketing*, *80*(2), 59–79. https://doi.org/10.1509/jm.15.0324

Klein, J., & Dawar, N. (2004). Corporate social responsibility and consumers' attributions and brand evaluations in a product–harm crisis. *International Journal of Research in Marketing*, *21*(3), 203–217. https://doi.org/https://doi.org/10.1016/j.ijresmar.2003.12.003

Kountouri, I., Manousakis, E., & Tsekrekos, A. E. (2019). Latent semantic analysis of corporate social responsibility reports (with an application to Hellenic firms). *International Journal of Disclosure and Governance*, *16*, 1–19. https://doi.org/10.1057/s41310-018-0053-z

Lee, K., Oh, W.-Y., & Kim, N. (2013). Social Media for Socially Responsible Firms: Analysis of Fortune 500's Twitter Profiles and their CSR/CSIR Ratings. *Journal of Business Ethics*, *118*, 791–806. https://doi.org/10.1007/s10551-013-1961-2

Lee, S., & Cho, M. (2011). Social media use in a mobile broadband environment: Examination of determinants of Twitter and Facebook use. *International Journal of Mobile Marketing*, *6*(2), 71–87.

Li, L., McMurray, A., Xue, J., Liu, Z., & Sy, M. (2018). Industry-wide corporate fraud: The truth behind the Volkswagen scandal. *Journal of Cleaner Production*, *172*, 3167–3175. https://doi.org/10.1016/j.jclepro.2017.11.051

Liao, P.-C., Liao, J.-Q., Wu, G., Wu, C.-L., Zhang, X.-L., & Ma, M.-C. (2018). Comparing international contractors' CSR communication patterns: A semantic analysis. *Journal of Cleaner Production*, *203*, 353–366. https://doi.org/10.1016/j.jclepro.2018.08.218

Lock, I., & Araujo, T. (2020). Visualizing the triple bottom line: A large-scale automated visual content analysis of European corporations' website and social media images. *Corporate Social Responsibility and Environmental Management*, *27*(6), 2631–2641. https://doi.org/10.1002/csr.1988

Lock, I., & Seele, P. (2015). Quantitative content analysis as a method for business ethics research. *Business Ethics: A European Review*, *24*(S1), S24–S40. https://doi.org/10.1111/beer.12095

Lock, I., & Seele, P. (2016). The credibility of CSR (corporate social responsibility) reports in Europe. Evidence from a quantitative content analysis in 11 countries. *Journal of Cleaner Production*, *122*, 186–200. https://doi.org/10.1016/j.jclepro.2016.02.060

Loviscek, V. (2020). Triple Bottom Line toward a Holistic Framework for Sustainability: A Systematic Review. *Revista de Administração Contemporânea*, *25*(3), e200017. https://doi.org/10.1590/1982-7849rac2021200017.en

Mann, M., Byun, S.-E., & Ginder, W. (2021). B Corps' Social Media Communications during the COVID-19 Pandemic: Through the Lens of the Triple Bottom Line. *Sustainability*, *13*(17), 9634. https://doi.org/10.3390/su13179634

Marcelino-Sádaba, S., González-Jaen, L. F., & Pérez-Ezcurdia, A. (2015). Using project management as a way to sustainability. From a comprehensive review to a framework definition. *Journal of Cleaner Production*, *99*, 1–16. https://doi.org/10.1016/j.jclepro.2015.03.020

Mayyas, A., Qattawi, A., Omar, M., & Shan, D. (2012). Design for sustainability in automotive industry: A comprehensive review. *Renewable and sustainable energy reviews*, *16*(4), 1845–1862. https://doi.org/10.1016/j.rser.2012.01.012

Miles, M. P., & Covin, J. G. (2000). Environmental marketing: A source of reputational, competitive, and financial advantage. *Journal of Business Ethics*, *23*(3), 299–311.

Miller, S. R., Eden, L., & Li, D. (2020). CSR Reputation and Firm Performance: A Dynamic Approach. *Journal of Business Ethics*, *163*(3), 619–636. https://doi.org/10.1007/s10551-018-4057-1

Minton, E., Lee, C., Orth, U., Kim, C.-H., & Kahle, L. (2012). Sustainable Marketing and Social Media: A Cross-Country Analysis of Motives for Sustainable Behaviors. *Journal of Advertising*, *41*(4), 69–84. https://doi.org/10.1080/00913367.2012.10672458

Mish, J., & Scammon, D. L. (2010). Principle-Based Stakeholder Marketing: Insights from Private Triple-Bottom-Line Firms. *Journal of Public Policy & Marketing*, *29*(1), 12–26. https://doi.org/10.1509/jppm.29.1.12

Mishra, S., & Modi, S. B. (2016). Corporate Social Responsibility and Shareholder Wealth: The Role of Marketing Capability. *Journal of Marketing*, *80*(1), 26–46. https://doi.org/10.1509/jm.15.0013

Porter, M. E., & Kramer, M. R. (2006). The link between competitive advantage and corporate social responsibility. *Harvard Business Review*, *84*(12), 78–92.

Raghubir, P., Roberts, J., Lemon, K. N., & Winer, R. S. (2010). Why, When, and How Should the Effect of Marketing Be Measured? A Stakeholder Perspective for Corporate Social Responsibility Metrics. *Journal of Public Policy & Marketing*, *29*(1), 66–77. http://www.jstor.org/stable/20798396

Richardson, A. J., & Welker, M. (2001). Social disclosure, financial disclosure and the cost of equity capital. *Accounting, Organizations and Society*, *26*(7-8), 597–616.

Russo-Spena, T., Tregua, M., & De Chiara, A. (2018). Trends and Drivers in CSR Disclosure: A Focus on Reporting Practices in the Automotive Industry. *Journal of Business Ethics*, *151*, 563–578. https://doi.org/10.1007/s10551-016-3235-2

Sanchez-Chaparro, T., Soler-Vicen, M. A., & Gomez-Frías, V. (2022). Be good and look good: Communicating the triple bottom line through corporate websites. *Journal of Business Research*, *144*, 136–145. https://doi.org/10.1016/j.jbusres.2022.01.089

Schaltegger, S., & Burritt, R. (2018). Business Cases and Corporate Engagement with Sustainability: Differentiating Ethical Motivations. *Journal of Business Ethics*, *147*, 241–259. https://doi.org/10.1007/s10551-015-2938-0

Shinkle, G. A., & Spencer, J. W. (2012). The social construction of global corporate citizenship: Sustainability reports of automotive corporations. *Journal of World Business*, *47*(1), 123–133. https://doi.org/10.1016/j.jwb.2011.02.003

Siano, A., Vollero, A., Conte, F., & Amabile, S. (2017). "More than words": Expanding the taxonomy of greenwashing after the Volkswagen scandal. *Journal of Business Research*, *71*, 27–37. https://doi.org/10.1016/j.jbusres.2016.11.002

Sukitsch, M., Engert, S., & Baumgartner, R. J. (2015). The implementation of corporate sustainability in the European automotive industry: An analysis of sustainability reports. *Sustainability*, *7*(9), 11504–11531.

United Nations. (2015). Transforming our World: The 2030 Agenda for Sustainable Development. *Department of Economic and Social Affairs*. https://sdgs.un.org/publications/transforming-our-world-2030-agenda-sustainable-development-17981

United Nations Global Compact, UC Berkley School of Law, & Linklaters LLP. (2019). *Guide for General Counsel om Corporate Sustainability Version 2.0*.

Wolff, S., Brönner, M., Held, M., & Lienkamp, M. (2020). Transforming automotive companies into sustainability leaders: A concept for managing current challenges. *Journal of Cleaner Production*, *276*, 124179.

Junior Management Science 9(1) (2024) 1140-1177

Junior Management Science

www.jums.academy
ISSN: 2942-1861

Flipping the Switch – The Role of Activity Load in Temporal Acquisition Patterns of Acquiring Firms

Frédéric Herold

University of St.Gallen

Abstract

This study presents evidence on the effect of a firm's activity load from acquisitions on its temporal acquisition pattern. Exploiting a panel of the 300 largest Fortune Global 500 firms over the 1990-2010 period, I use a hybrid logit model in which I regress momentum on activity load. I find that increases in the activity load from acquisitions, on average, reduce a firm's likelihood to maintain acquisition momentum. That is, the increase in acquisition activity created by acquisition routines and cognitive maps of managers translates into a higher activity load until firms face a situation of information overload. Rational acquirers neutralize this pressure by reducing their acquisition volume which, in turn, decreases the activity load burden. Moreover, my results reveal that acquirers can switch from targets in a higher-complexity target firm category to targets in a lower-complexity target firm category to reduce their activity load burden while maintaining overall momentum. Yet, I obtain ambiguous results when examining heterogeneity in acquirer responses arising from differences in absorptive capacity. With these findings, my study adds to prior literature on acquisition patterns, strategic momentum, and the interplay between a firm's activity load and absorptive capacity.

Keywords: activity load; M&A; momentum; switching behavior; temporal acquisition patterns

1. Introduction

1.1. Motivational Background

Mergers and acquisitions (M&A) are a phenomenon that has received substantial attention from scholars and practitioners alike. This is in part due to the plethora of reasons why firms engage in M&A, which range from entering new markets and accessing new resources (e.g., Karim & Mitchell, 2000; Lee & Lieberman, 2010) to creating economies of scale/scope (e.g., Biggadike, 1979; Lee & Lieberman, 2010). Also, this is in part due to the sobering performance of M&As, which on average is non-existent or even negative (King et

I gratefully acknowledge the support and guidance of my first supervisor, Prof. Dr. Dr. Tomi Laamanen, over the course of this research project. Specifically, I would like to recognize his generosity in proposing this topic and engaging in numerous valuable discussions with me that laid the theoretical and empirical foundation of this study. Furthermore, I would like to express my sincere gratitude to my second supervisor, Prof. Dr. Xena Welch Guerra, for our fruitful discussions in which she provided me with critical feedback. Also, I would like to thank Prof. Dr. Johannes Luger for advising me on the use of empirical methods and providing the raw data on acquisitions, human-coded TMT member titles, and business-segment fundamentals without which my empirical analysis would not have been possible in its present form. Finally, I would like to extend my gratitude to Nicholas Herold and Ralf Herold for their thoughtful comments and invaluable suggestions that substantially contributed to the quality of this study. Beyond these decisive contributions, I would like to acknowledge that this study builds on and has substantially benefitted from a draft version of a paper circulating under the title "Temporal Dynamics in Acquisition Behavior: The Effects of Activity Load on Strategic Momentum" written by Yuval Deutsch, Thomas Keil, Tomi Laamanen, and Markku Maula. Furthermore, this thesis has substantially benefitted from the methodological advice and code provided by Trenton D. Mize and colleagues (2019). The authors generously provided STATA script files, which I adapted to my research setting, on the personal website of Trenton D. Mize under https://www.trentonmize.com/software/mecompare. Please refer to their paper "A General Framework for Comparing Predictions and Marginal Effects Across Models" for more details. All tables and figures presented in this study are own visualizations unless indicated otherwise.

DOI: https://doi.org/10.5282/jums/v9i1pp1140-1177

al., 2004). Even so, the fascination for and practical importance of M&A remain unbroken as evidenced by a global M&A transaction value of USD 3.9tn in 2019 alone (Roumeliotis & Barbaglia, 2019).

While scholars in the field of M&A have predominantly studied acquisitions as singular events (S. Chatterjee, 2009), they have recently also shown a growing interest in acquisition streams (Shi et al., 2012). Scholars who take this perspective view acquisitions as sequences of interrelated strategic action that can be actively managed by acquirers rather than as isolated, exclusively opportunity-driven events (Schipper & Thompson, 1983; Shi et al., 2012). This interest manifested itself in a growing body of literature that has focused on identifying distinct acquisition patterns (e.g., Shi & Prescott, 2011), their observable properties (e.g., Laamanen & Keil, 2008; Vermeulen & Barkema, 2002), and their firm-level performance implications for acquirers (e.g., Laamanen & Keil, 2008; Shi & Prescott, 2012; Vermeulen & Barkema, 2002).

However, much less attention has been attributed to factors that explain why firms *deviate* from their acquisition patterns. That is, while scholars have examined how such patterns emerge and why acquirers do *follow* their patterns over time (e.g., Amburgey & Miner, 1992), compelling theoretical explanations for and empirical evidence on *how* and *why* firms would *deviate* from their established acquisition patterns has remained surprisingly scarce. This gap is puzzling for two reasons. First, past studies have argued and found that firms show acquisition behavior that is *inconsistent* with their past acquisition pattern (e.g., Vermeulen & Barkema, 2001), indicating that this phenomenon is much more than a merely theoretical question. Second, deviations from previously stable acquisition patterns are detrimental to acquirer performance (e.g., Ellis et al., 2011; Laamanen & Keil, 2008). Considering these implications, managers, M&A advisors, and scholars would benefit from better understanding the factors that cause acquirers to systematically deviate from their established acquisition patterns.

This thesis aims to do exactly that by studying a factor that is novel to acquisition pattern research: Activity load. Specifically, I argue that acquisition streams can create high levels of activity load that overstretch the managerial resources of acquirers and, thus, lead to a situation of 'overload' (Castellaneta & Zollo, 2015; Laamanen & Keil, 2008; Vermeulen & Barkema, 2002). This overload results in 'corporate indigestion' (Kusewitt, 1985), which induces firms to deviate from their stable acquisition patterns to alleviate the strains on their resources. That is, firms can respond in two ways. They can either *switch* to acquisitions that create a lower activity load (e.g., acquire targets that are *relatively* less complex to acquire) or *decelerate* their acquisition pace (i.e., acquire fewer targets without changing the type of target acquired). Both responses allow firms to reduce the activity load from acquisitions, albeit with different implications for their acquisition patterns. Which response prevails is a question that can be empirically answered by studying the effect of activity load on acquisition behavior.

1.2. Focus and Objectives of Thesis

This thesis seeks to bridge the research gap identified in section 1.1 by presenting empirical evidence on *how* and *why* acquiring firms deviate from their established acquisition patterns. Specifically, by investigating the role of activity load in this phenomenon, this thesis aims to answer the following research question:

***Research question**: How does the activity load from acquisitions cause acquiring firms to deviate from their established acquisition patterns?*

To approach this question, I break it down into three subquestions with different objectives as illustrated in Figure 1. Each subquestion and its objectives are briefly outlined in the following.

***Subquestion 1**: How does the volume of acquisitions in an acquisition stream affect a firm's acquisition behavior?*

With subquestion 1, I investigate how the *volume* of acquisitions in an acquisition stream affects the acquisition behavior of firms. I pursue two objectives with this. First, I seek to conceptually understand the core mechanism through which activity load causes firms to deviate from their established acquisition patterns. Since activity load is a complex construct, which consists of a volume *and* a complexity component, this objective is best achieved by isolating the construct's most intuitive-to-understand component: Acquisition *volume*. Thus, the relationship between an acquirer's acquisition *volume* and its acquisition behavior will serve as the baseline effect of activity load in this thesis. Second, building on this theoretical basis, I aim to empirically confirm the baseline relationship between acquisition volume and the acquisition behavior of firms. This would already provide a first answer to my research question and – most importantly – create the basis for subsequent analyses of cross-sectional variation in acquisition behavior.

***Subquestion 2**: How does a firm's change in acquisition behavior vary with the complexity level of acquisitions?*

Subsequently, with subquestion 2, I analyze cross-sectional variation in acquirer responses to activity load with respect to the *complexity* level of acquisitions. With this, I want to reach two objectives. First, from a theoretical standpoint, I seek to identify externally observable structural target firm attributes in literature that drive acquisition complexity and, thus, the activity load of acquirers. With this knowledge, I can go beyond the baseline effect of acquisition volume and theorize more nuanced behavioral responses of acquirers to changes in their activity load. Second, from an empirical standpoint, I aim to corroborate relevant complexity dimensions and the changes in acquisition behavior they induce. This allows me to explore the interaction of both activity load components and provide a more fine-grained answer to my research question.

***Subquestion 3**: Which firm-level factors moderate the relationship between activity load and acquisition behavior?*

Finally, with subquestion 3, I explore a second source of cross-sectional variation in acquisition behavior: Structural characteristics of acquirers. I hope to accomplish two objectives with this. First, as with subquestion 2, I aim to find

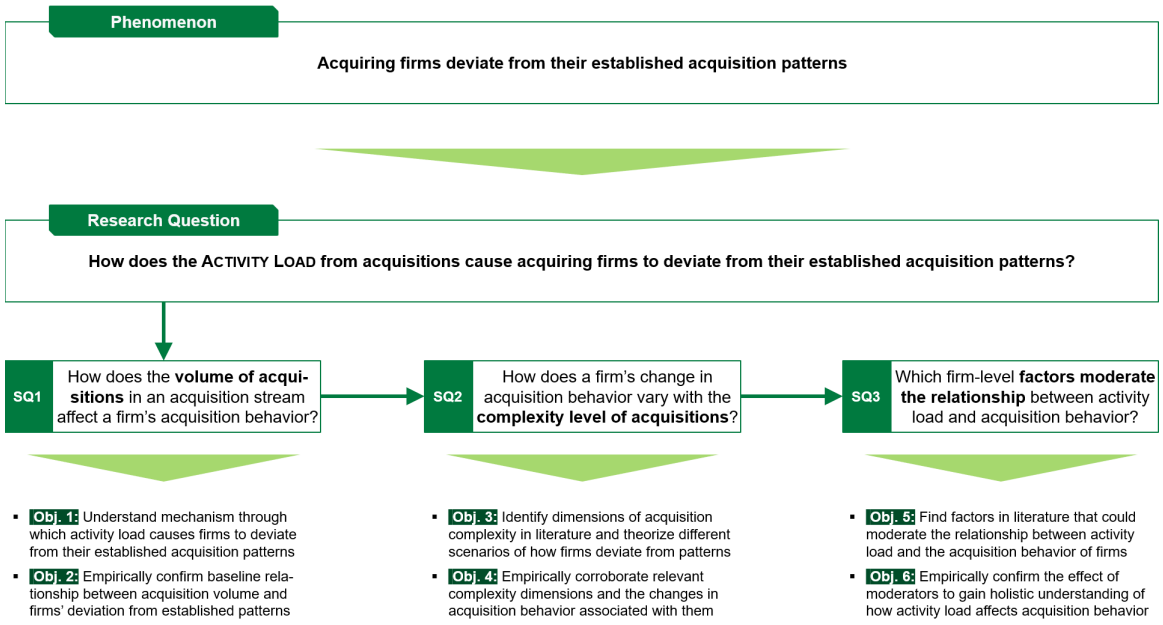


Figure 1: Focus and Objectives of Thesis

firm-level factors in literature that could moderate the relationship between activity load and the acquisition behavior of firms. This allows me to fine-tune my theoretical predictions of changes in acquisition behavior by putting the activity load from acquisitions in relation to a firm's capacity to absorb these acquisitions. Second, I seek to empirically confirm the effect of these moderators. Through this, I can corroborate predictions of earlier studies, which argued that the absorptive capacity of firms can alleviate the effects of activity load (e.g., Penrose, 1959). Most importantly, however, this allows me to gain a holistic understanding of how the activity load from acquisitions affects acquisition behavior of firms.

To derive meaningful insights from my empirical analyses, I made three scoping decisions that allow this thesis to answer the research question while complying with quality criteria used in business research (Bryman & Bell, 2011). First, by investigating a firm-level phenomenon, this thesis focuses on the individual firm as the unit of analysis. Second, my causal inference is based on panel data of acquiring firms and a time-series logit model. This approach accounts for the fact that acquisition patterns evolve over time (Laamanen & Keil, 2008; Shi & Prescott, 2011, 2012). Third, to ensure generalizability of empirical results, this thesis exploits a sample[1] that covers a 21-year acquisition period of acquirers listed in the Fortune Global 500 ranking, resulting in a sample with a long timeframe and broad industry scope. Although not without limitations[2], the chosen scope overall allows me to investigate the phenomenon of interest over a sufficiently long period and exploit a rich sample of firms which are highly acquisitive.

1.3. Definitions of Key Terms

By examining whether activity load stimulates acquiring firms to deviate from their established acquisition patterns, this thesis investigates a research question that builds on four key concepts: Acquisitions, temporal acquisition patterns, strategic momentum, and activity load. These key concepts must be explicitly defined to ensure clarity. Thus, I present their definitions hereafter.

Acquisition. In line with the Refinitiv Eikon M&A database (Refinitiv, n.d.), this thesis defines an acquisition as an economic transaction between an acquirer and the shareholders of a target that involves the transfer of ownership rights at the level of the ultimate parent. This definition allows me to include *all* known deals of an acquiring firm in my analysis, irrespectively of the ownership stake transferred. Through this, I obtain a realistic measure of a firm's activity load from acquisitions, which reflects *all* acquisition-related activities that consume firm resources.

Temporal acquisition patterns. Temporal acquisition patterns are sequences/programs of inter-related acquisitions which are directed at executing a firm's strategy (Laamanen & Keil, 2008; Schipper & Thompson, 1983), capturing the systematic acquisition behavior of firms over time. Literature has defined temporal acquisition patterns in terms of their (i) mathematical properties (i.e., acquisition frequency and variability thereof) (Laamanen & Keil, 2008) and (ii) externally observable structural target firm attributes (e.g., size, industry relatedness, or location) (Ellis et al., 2011; Hayward, 2002). I will use the latter definition (i.e., observable

[1] Please refer to sections 3.1 and 6.1 for an illustration of the rationale behind and limitations of the chosen sample, respectively.
[2] See section 6.1 for details.

target firm attributes) in this thesis to ensure comparability of my results with past studies of strategic momentum.

Strategic momentum. Following Amburgey and Miner (1992), I define strategic momentum as "[a firm's] tendency to maintain or expand the emphasis and direction of prior strategic actions in current strategic behavior" (p. 335). Applied to the acquisition context, momentum can help explain not only why, *in general*, acquirers are more likely to engage in subsequent acquisitions if they acquired firms in the past but also why prior experience in acquisitions *of a specific type* (e.g., cross-border acquisitions or acquisitions in unrelated industries) increases the propensity of acquirers to keep acquiring targets of the same type (Amburgey & Miner, 1992).

Activity load. A firm's activity load corresponds to the level of simultaneous activity in which that firm engages at a given time (Castellaneta & Zollo, 2015). It is driven by the number of parallel activities and their respective complexity (Castellaneta & Zollo, 2015). In the context of acquisitions, high levels of activity load can create an overload situation for acquiring firms (Castellaneta & Zollo, 2015) that could explain deviations from established acquisition patterns.

1.4. Structure of Thesis

This thesis is structured into six sections. Section 1 provides an entry point into the topic by explicating the underlying motivation of this thesis and delineating the research question, which I break down into three subquestions. In section 2, I establish the theoretical foundation of this thesis, based on which I develop my hypotheses. For this, I review literature on (a) temporal acquisition patterns, (b) strategic momentum, and (c) activity load/absorptive capacity to gain a complete understanding of the phenomenon of interest and correctly position this thesis in extant literature. To ensure replicability of my findings, section 3 illustrates the methodological approach used in this thesis, describing all data collection and modification steps, the variables, and the estimation strategy. Subsequently, my descriptive statistics, empirical results of the hybrid logit model, and robustness tests are presented in section 4. Section 5 then reconciles my findings with prior literature, reviewing contributions of this study to academia and practice, discussing limitations of this thesis, and highlighting avenues for future research. Finally, my conclusions are set forth in section 6. Figure 2 summarizes my approach.

2. Theoretical Background

To illustrate my phenomenon of interest and correctly position this thesis in extant research, I begin this section with a brief review of the temporal perspective used in acquisition research, in general, and temporal acquisition patterns, in particular. Subsequently, I introduce the key concepts of (a) strategic momentum, (b) activity load, and (c) absorptive capacity that will form the theoretical basis of this thesis. Finally, I delineate the mechanism through which activity load induces acquiring firms to deviate from their established acquisition patterns and formally develop my hypotheses, which will be tested in section 4.

2.1. Temporal Acquisition Patterns

Over the last three decades, research on acquisitions and related fields, such as alliances, has increasingly adopted a *temporal perspective*, primarily investigating the performance effects of time-related constructs and phenomena (Shi et al., 2012). This rise in popularity has translated into a broad range of research questions and methodological approaches used to study these topics (Shi et al., 2012). For example, extant acquisition research has investigated the timing of acquisitions relative to environmental factors, such as industry M&A waves (e.g., McNamara et al., 2008), acquisition moves of competitors (e.g., Carow et al., 2004), non-M&A corporate development initiatives, such as alliances (e.g., Shi & Prescott, 2011, 2012), and previous acquisitions of acquiring firms (e.g., Haleblian & Finkelstein, 1999; Hayward, 2002), among others. This breadth of research contexts exemplifies the versatility of the temporal perspective and, thus, helps explain its frequent adoption in acquisition research[3].

One growing literature stream which extensively uses this perspective has studied *temporal acquisition patterns* of firms. These patterns are sequences of interrelated acquisitions which are directed at executing a firm's strategy (Laamanen & Keil, 2008; Schipper & Thompson, 1983). Scholars have studied this phenomenon in different ways. For instance, in an exploratory study of acquisition patterns, Shi and Prescott (2011) have developed a pattern taxonomy, deriving seven distinct patterns and grouping those into three clusters (i.e., predictable patterns, unpredictable patterns, no acquisition activity) with different performance implications. Other studies go one step further and either (a) break down these acquisition patterns into numerical subcomponents, such as the *rate* (or *frequency*) of acquisitions within a given period and the *variability* in this acquisition rate over time (or *rhythm*) (e.g., Laamanen & Keil, 2008; Shi & Prescott, 2012; Vermeulen & Barkema, 2002), or (b) identify such patterns based on externally observable target firm attributes, such as target size[4] (e.g., Ellis et al., 2011), industry relatedness (e.g., Hayward, 2002), or geographic location (e.g., Vermeulen & Barkema, 2002). These distinctions allowed scholars to develop a more fine-grained understanding of acquisition patterns and, thus, have frequently been used in recent acquisition pattern research.

Thus far, extant research on these patterns has intensively investigated their performance effects for acquiring firms,

[3] Relevant streams include organizational learning (e.g., Hayward, 2002), internationalization processes (e.g., Vermeulen & Barkema, 2002), and strategic change (e.g., Klarner & Raisch, 2013), to name a few. See Shi et al. (2012) for a comprehensive review.

[4] Although Ellis et al. (2011) did not use the term acquisition pattern, they de facto study acquisition patterns by analyzing how size-specific acquisition experience, which acquirers gain throughout their acquisition history (i.e., acquisition pattern), is related to performance.

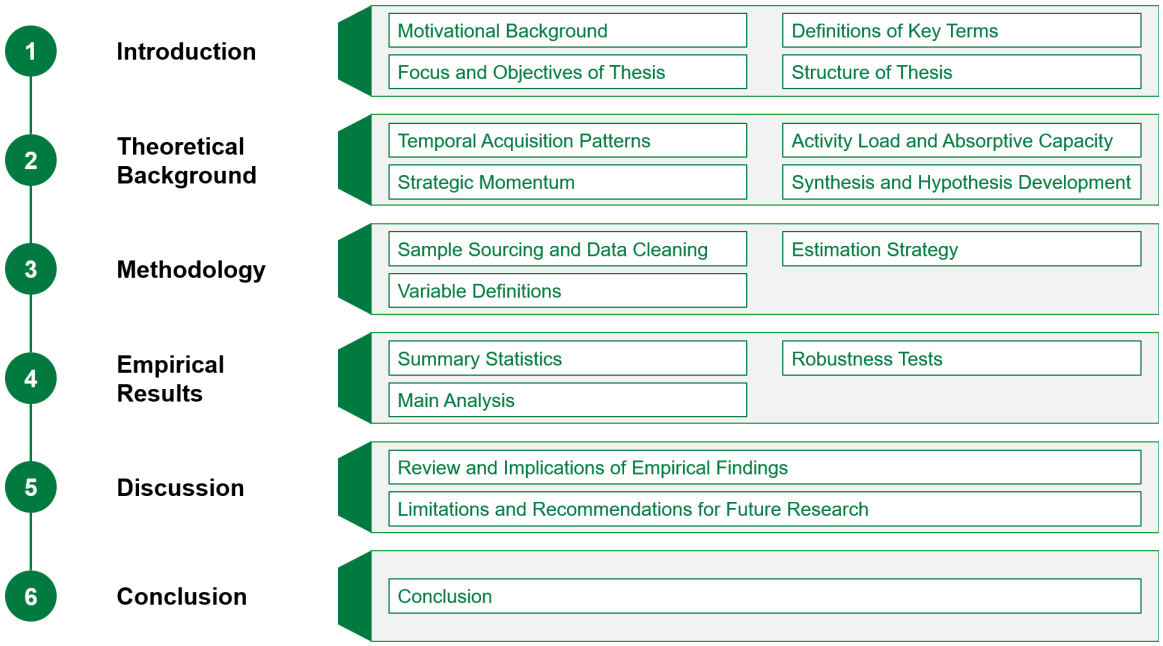


Figure 2: Structure of Thesis

highlighting the practical relevance of this phenomenon. Specifically, most studies[5] find that *stable* acquisition patterns are associated with *positive* performance outcomes for acquiring firms (e.g., Hayward, 2002; Shi & Prescott, 2011), whereas *irregular* patterns or *deviations* from stable patterns *adversely* affect the performance of these firms (e.g., Ellis et al., 2011; Laamanen & Keil, 2008; Vermeulen & Barkema, 2002). To explain these results, studies have primarily drawn on organizational learning theory and Penrosian resource-based logic. That is, acquiring firms perform better (worse) if they acquire targets in predictable (irregular) intervals that (do not) allow firms to infer correct learnings from prior acquisition experience and codify these in routines (Hayward, 2002; Laamanen & Keil, 2008; Shi & Prescott, 2011), refrain from applying (attempt to transfer) the same acquisition routines to dissimilar target firms (Ellis et al., 2011), and more evenly use (abruptly exceed) the limited capacity of their managers (Laamanen & Keil, 2008; Shi & Prescott, 2011). Considering these striking differences in acquirer performance, one would thus expect scholars to have dug deeper into factors that explain the emergence of and deviation from stable acquisition patterns.

However, as shown in Figure 3 our understanding of the antecedents of temporal acquisition patterns remains rather limited, indicating a substantial gap in acquisition research. This might, at first glance, seem surprising given the myriad of factors that scholars found to influence acquisition behavior of firms. Yet, a closer look at the literature on antecedents of acquisition behavior reveals that this discrepancy likely arises from two different research questions to which scholars have devoted unequal levels of attention: (a) What makes some firms acquire more than others (i.e., differences in acquisitiveness) and (b) what causes the emergence of and deviation from acquisition patterns (i.e., temporal acquisition patterns). Specifically, scholars have extensively focused on the first research question, discussing the role of a vast array of antecedents of acquisition behavior. These antecedents cover factors on the individual/team level, such as CEO characteristics (e.g., A. Chatterjee & Hambrick, 2007; Gamache et al., 2015; Malmendier & Tate, 2008; Seo et al., 2015) and changes within the acquiring firm's top management team (TMT) (e.g., Shi et al., 2017); acquiring firm level, such as a firm's ambition level (e.g., Haleblian et al., 2017; Kim et al., 2015) and past growth history (e.g., Kim et al., 2011); and environmental level, such as imitation of other firms' acquisition behavior (e.g., Baum et al., 2000; Haunschild, 1993; Haunschild & Beckman, 1998) and the culture-/industry-specific context of acquiring firms (e.g., Bertrand et al., 2019), for instance. However, only few studies have attempted to answer the second research question. These studies (e.g., Amburgey & Miner, 1992) have found that past acquisition behavior helps explain how temporal acquisition patterns emerge and why firms systematically adhere to their patterns over time, complementing prior research on acquisitiveness. Yet, it appears that no study that focused on antecedents of acquisition behavior has investigated possible factors that cause acquiring firms to *systematically deviate*[6] from their established acquisition patterns.

[5] Shi and Prescott (2012) argue that firms need to strike a balance between the high internal coordination costs of unpredictable patterns and the flexibility to engage in sudden acquisition opportunities, predicting an inverted u-shape relationship between temporal rhythm and acquirer performance. However, using a single-industry sample, they acknowledge the limited generalizability of their results.

[6] Systematic deviations are permanent shifts in acquisition behavior (e.g.,

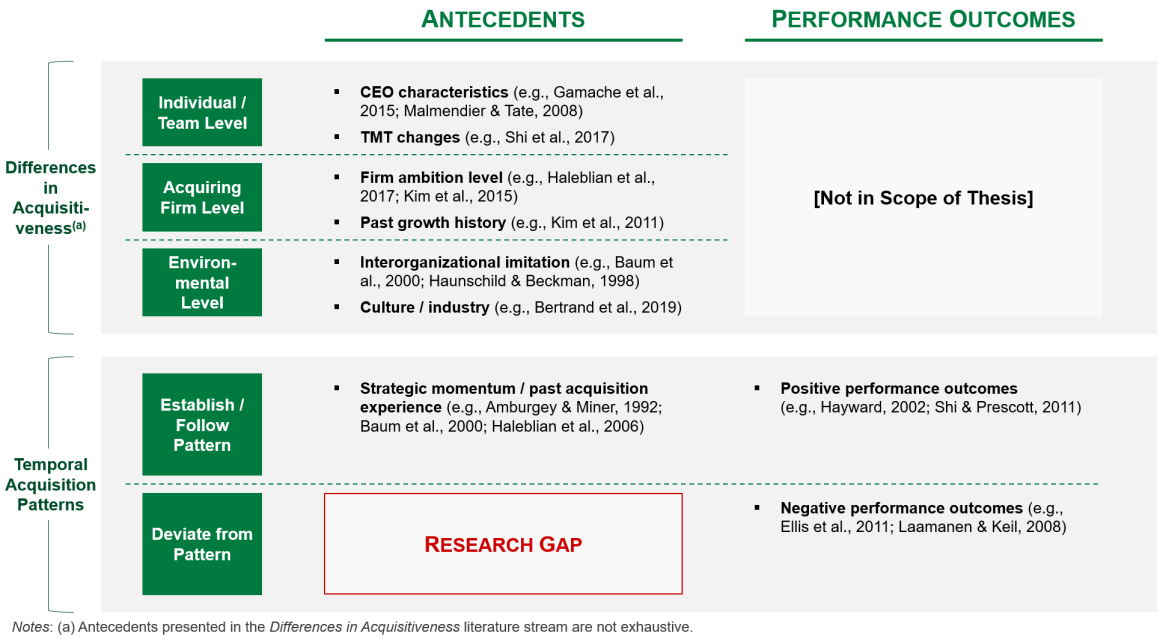


Figure 3: Research Gap in Literature on Antecedents of Temporal Acquisition Patterns

This thesis thus seeks to bridge this gap by introducing activity load as a new concept to acquisition pattern research.

2.2. Strategic Momentum

Scholars in the field of organizational science introduced the concept of *strategic momentum* four decades ago. Defining "momentum as the tendency [of firms] to maintain or expand the . . . direction of prior strategic actions in current strategic behavior" (Amburgey & Miner, 1992, p. 335), early studies of strategic momentum primarily investigated this concept in the context of organizational change (e.g., Kelly & Amburgey, 1991; Miller & Friesen, 1980) and innovation (e.g., Miller & Friesen, 1982), obtaining remarkably consistent results. For instance, scholars of organizational change found not only that firms are more likely to implement changes that are consistent with their adopted strategic and structural orientation (Miller & Friesen, 1980) but also that firms which experience strategic changes of one type (e.g., a change from a generalist strategic orientation to a specialized orientation) keep making strategic changes of the same type (Kelly & Amburgey, 1991). Likewise, studies of innovation patterns of firms showed that past innovation practices evolve in the same direction (Miller & Friesen, 1982), corroborating that actions of one type increase the likelihood of subsequent similar actions (Haleblian et al., 2006) and that the concept of momentum can hold in multiple research settings.

Building on these early studies, Amburgey and Miner (1992) applied the concept of momentum to the context of acquisitions, showing that momentum can help explain acquisition behavior, in general, and the emergence of acquisition patterns, in particular. For this, they distinguished between three different types of momentum: Repetitive momentum (the tendency to repeat past strategic actions), positional momentum (the tendency to maintain or extend current strategic positions), and contextual momentum (the influence of contextual features on strategic actions). Using a multivariate point process model, the authors found empirical support for the existence of repetitive and contextual momentum. Specifically, they observed that past experience in acquisitions of one type, such as horizontal acquisitions, increases the propensity of acquirers to engage in subsequent acquisitions of the same type (repetitive momentum), while acquirers with a decentralized structure are more likely to pursue diversification acquisitions (contextual momentum). Although Amburgey and Miner (1992) could only partially confirm the existence of positional momentum[7], their study was the first to show that firms follow acquisition patterns that are consistent with past experiences, expanding prior research on strategic momentum.

To explain their results, Amburgey and Miner (1992) resorted to organizational learning theory and theories of managerial cognition. Specifically, they argued that, by acquiring

an acquiring firm that previously acquired large targets now switches to small targets). As I will explain in sections 2.3 and 2.4, such a pattern shift involves substantial costs and, thus, can likely not be explained by antecedents of acquisition behavior that primarily explain differences in the number and volume of acquisitions. If firms are influenced by these factors, it thus appears more likely that firms simply adjust the number of deals and deal volume instead of changing their patterns.

[7] Firms with a diversified position in multiple product markets tend to expand their current position through further product market diversification acquisitions. However, conglomerate firms do not seem to engage in further conglomerate acquisitions (Amburgey & Miner, 1992).

targets of a certain type, acquirers accumulate type-specific acquisition experience, which (a) leads to the formation of type-specific acquisition routines and (b) positions these targets more centrally in the cognitive maps of managers. Both concepts explain the creation of acquisition momentum, although for different reasons. *Routines*, on the one hand, "are programs of action that . . . [formalize an organization's] experience . . . with a particular task" (Haleblian et al., 2006, p. 358), constituting a key element of organizations (Cyert & March, 1963; Nelson & Winter, 1982). They are formed through repeated actions, which allow organizations to build up competence and confidence in executing these actions, creating an incentive to repeat learned behaviors (Collins et al., 2009) and, thus, momentum (Amburgey & Miner, 1992). In other words, through routines, "an organization undertakes . . . activities . . . because it knows how to [execute these activities]"[8] (Amburgey & Miner, 1992, p. 336). In the context of acquisitions, routines encompass "templates for selecting . . . [specific] targets or [post-integration] guidelines" (Haleblian et al., 2006, p. 358) and therefore help explain type-specific acquisition momentum (Amburgey & Miner, 1992). *Cognitive maps*, on the other hand, refer to the shared mindsets of managers. They are shaped by experiences in certain strategic actions, such as acquisitions of targets of a specific type, inducing a preference for these actions (Amburgey & Miner, 1992). That is, once managers build up experience with targets of a specific type, these targets "[take] a more central role in [their] . . . cognitive [maps and, thus, are] . . . more likely to be seen as an appropriate [strategic action]" (Amburgey & Miner, 1992, p. 337). Both concepts therefore corroborate that acquisitions are subject to inertial pressures, promoting acquisitions of familiar target types (i.e., more exploitation) while reducing an acquirer's likelihood to buy less-known target types (i.e., less exploration) (Collins et al., 2009).

As presented in Table 1, research on acquisition patterns has frequently adopted the reasoning of Amburgey and Miner (1992) and extensively studied momentum as a firm-level explanation of acquisition behavior, substantiating the repetitive momentum hypothesis. One set of studies, for example, explored *overall* acquisition momentum, finding that prior acquisition experience, *in general*, increases the likelihood of subsequent acquisitions (Haleblian et al., 2006). Other studies chose a more fine-grained perspective by differentiating between externally observable target firm attributes. In a study of local expansion patterns of firms, for example, Baum et al. (2000) showed that firms are more likely to acquire targets that have attributes *similar* to those of previously acquired targets. Likewise, Collins et al. (2009) found that prior experience in cross-border acquisitions increases the likelihood of firms to engage in further acquisitions of that type, corroborating that repetitive momentum holds in a cross-border acquisition context. Both sets of studies therefore found a *positive* relationship between prior acquisition experience and subsequent acquisition activity – a result consistent with predictions of momentum theory.

However, this consensus has been challenged recently. Albeit in the context of organizational change, Beck et al. (2008) found that firms exhibit behavior *opposite* to the repetitive momentum hypothesis, providing both theoretical and methodological explanations for their observation. From a theoretical viewpoint, the authors argued that prior change enables firms "to refine the content of organizational procedures . . . [and] routines that govern change processes" (p. 428) which reduces the need for subsequent change, although they did not directly measure this mechanism. Methodologically (and more dramatically), Beck et al. (2008) showed that accounting for unobserved time-invariant heterogeneity on the firm level *reverses* the direction of the momentum effect (i.e., prior change triggers a *deceleration* in further change of the same type, not an acceleration). They confirmed their observations with three datasets in different research settings, indicating that past studies of momentum, which mostly use random effects models, have likely reported biased results. Moreover, Vermeulen and Barkema (2001) presented findings in a cross-border acquisition context that are similar to those of Beck et al. (2008). That is, after including firm fixed effects in their logit model, Vermeulen and Barkema (2001) showed that prior cross-border acquisitions trigger a *deceleration* in further cross-border acquisition activity[9], demonstrating that the results of Beck et al. (2008) are equally relevant for scholars who study acquisition patterns. Both studies thus challenge the repetitive momentum hypothesis and reveal a major gap in momentum research that, I argue, can (at least partially) be closed by introducing the concept of activity load to momentum research.

2.3. Activity Load and Absorptive Capacity

Like strategic momentum, the concepts of *activity load* and *absorptive capacity* are not new to management research. Specifically, both date back to early studies of organizational behavior (Simon, 1945) and growth (Penrose, 1959), playing a particularly crucial role in Penrose's *The Theory of the Growth of the Firm*. In her study, Penrose (1959) has argued that the capacity of managers to engage in activities of firms (i.e., their absorptive capacity) is limited[10], essentially referring to Simon's (1945) idea of 'bounded rationality' according to which managers have a finite capacity to process information[11]. To grow, however, a firm needs its managers to en-

[8] Seminal studies of routines (e.g., Levitt & March, 1988; Nelson & Winter, 1982) emphasize that this "learning process is largely independent of the performance outcomes of prior experiences" (Haleblian et al., 2006, p. 357) since positive outcomes are interpreted as evidence of successful actions while negative outcomes are attributed to external circumstances or poor execution of routines (Amburgey & Miner, 1992).

[9] The authors report the increase in cross-border greenfield investments as an inverse proxy for the deceleration in cross-border acquisitions.

[10] That is, "there is plainly a physical maximum to the number of things any individual or group of individuals can do" (Penrose, 1959, p. 41).

[11] As Penrose (1959, p. 15) notes, "the general view . . . set forth here does not differ fundamentally from the concepts . . . of Simon [(1945)]".

Table 1: Overview of Studies of Strategic Momentum

Study	Literature Stream	Key Finding	Repetitive Momentum Hypothesis Confirmed
Miller and Friesen (1980)	Organizational Change	Organizations generally implement changes that are *consistent* with their adopted strategy and structure. However, periods of substantial *reversal* can occur if excesses / problems become dominant.	Partially[(a)]
Miller and Friesen (1982)	Innovation	Past innovation practices of proactive innovators (entrepreneurial firms) and reactive innovators (conservative firms) *keep evolving in the same direction*.	Yes
Kelly and Amburgey (1991)	Organizational Change	Firms which experience strategic changes of one type are *more likely* to pursue similar changes in the future.	Yes
Amburgey and Miner (1992)	Acquisition Patterns	Past experience in mergers of one type *increases* the number of subsequent mergers of the same type.	Yes
Baum et al. (2000)	Acquisition Patterns	Building on prior experience, firms are *more likely* to acquire targets which are similar to previously acquired targets than targets which are dissimilar.	Yes
Vermeulen and Barkema (2001)	Acquisition Patterns	Increases in cross-border acquisitions *decrease* the likelihood of engaging in subsequent cross-border acquisitions.	No
Haleblian et al. (2006)	Acquisition Patterns	Prior acquisition experience, positive performance feedback, and their interaction all *increase* the likelihood of engaging in subsequent acquisitions.	Yes
Beck et al. (2008)	Organizational Change	Prior changes trigger a *deceleration* of further changes of the same type.	No
Collins et al. (2009)	Acquisition Patterns	Prior acquisition experience in cross-border acquisitions *increases the likelihood* of engaging in subsequent cross-border acquisitions.	Yes

Notes: (a) This study finds an oscillation pattern. That is, periods of momentum alternate with periods in which a firm's direction of change reverses.

gage in growth-promoting activities[12] that consume their limited capacity (i.e., to shoulder the load imposed by these activities). By synthesizing these two premises, Penrose (1959) concludes that the growth of a firm depends on how much of the managerial capacity *available* to the firm is *consumed* by growth-promoting activities, with the *available* managerial capacity setting the upper growth limit. In fact, the interplay between the *availability* and *consumption* of managerial capacity is so central to her theory that Penrose (1959) even calls it the 'fundamental ratio'[13]. The concepts of activity load and absorptive capacity thus represent two sides of the same coin, adding to our understanding of organizations only in combination with each other.

Later studies have extended the early work of Penrose (1959), defining *activity load* as the level of simultaneous activity in which a firm engages at a given time (Castellaneta & Zollo, 2015) and distinguishing between two different drivers of a firm's activity load. This new distinction led to the emergence of two complementary perspectives in activity load research (Castellaneta & Zollo, 2015). One set of studies (e.g., Barkema & Schijven, 2008; Castellaneta & Zollo, 2015), for instance, has considered the concept of activity load from a purely quantitative view, the *computation* perspective, equating a firm's activity load with the *number* of simultaneous activities in which its managers engage

[12] Growth-promoting activities refer to what Penrose (1959) called "the creation and execution of plans for expansion" (p. 46).

[13] Penrose (1959) defines this ratio as "the ratio between the managerial services available [emphasis added] for expansion and the managerial services required [emphasis added] per dollar of expansion" (p. 175).

(first driver) (Castellaneta & Zollo, 2015). That is, firms bear a higher activity load if they, *ceteris paribus*, engage in a higher number of parallel activities. However, although this purely quantitative perspective offers compelling advantages, such as ease of empirical measurement, to scholars, it abstracts from the fact that activities can vary with respect to the amount of time and effort managers need to invest in them (Castellaneta & Zollo, 2015) and, thus, how much managerial capacity they consume. To bridge this gap, a second set of studies (e.g., Daft & Weick, 1984; Zorn et al., 2019) has adopted a different view of the activity load concept – the *interpretation* perspective. According to this view, the activity load of a firm varies with the level of difficulty its managers face in interpreting the information associated with the activities they pursue, depending on the *level of complexity*[14] of these activities (second driver) (Castellaneta & Zollo, 2015). In other words, firms bear a higher activity load if they, *ceteris paribus*, pursue activities that are more complex. Taken together, both perspectives show that activity load is a multifaceted concept that can only be fully understood once broken down into its components of volume and complexity.

Irrespectively of the activity load perspective adopted by these studies, they have unanimously argued that high levels of activity load can create problems for firms. That is, once the activity load of firms exceeds their *absorptive capacity*, which equals the maximum level of activity that a firm can simultaneously absorb, these firms face a situation of *information overload* (Castellaneta & Zollo, 2015). In such a situation, firms are overwhelmed by the amount and/or complexity of the information they need to digest (Castellaneta & Zollo, 2015). This makes them less able to infer correct learnings from past experience (Haleblian & Finkelstein, 1999; Hayward, 2002) due to 'time compression diseconomies' (Dierickx & Cool, 1989), reduces their capacity to absorb further activities (Vermeulen & Barkema, 2002), and – especially in the acquisition context – causes 'corporate indigestion' (Kusewitt, 1985). Building on this logic, scholars in the field of M&A have argued that an overload situation is particularly likely in the context of acquisitions (Castellaneta & Zollo, 2015), drawing on both activity load perspectives. On the one hand, "[acquiring] firms frequently engage in multiple acquisitions to execute their strategy" (Laamanen & Keil, 2008, p. 663), often managing a high volume of activities in parallel (computation perspective) (Castellaneta & Zollo, 2015). On the other hand, acquisitions themselves are strategic activities and, thus, "particularly complex ..., making information overload more likely" (interpretation perspective) (Castellaneta & Zollo, 2015, p. 142). In sum, research on activity load has agreed on the fact that overload can create considerable problems for firms and highlighted the relevance of information overload in the context of acquisitions.

Scholars generally agree that a firm's absorptive capacity cannot be easily expanded in the short run[15] (Penrose, 1959; Shaver, 2006). Yet, research has shown that some firm-level factors *can* alleviate the strains imposed by high levels of activity load. For this, studies have differentiated between three groups of factors. First, one stream of research (e.g., Kusewitt, 1985; Laamanen & Keil, 2008; Penrose, 1959) has stressed the importance of structural features of acquiring firms, such as size or organizational structure, in alleviating the burden imposed by high levels of activity load. Following Penrosian resource-based logic, these studies have argued that a larger size allows acquirers to not only access a larger pool of managerial resources but also benefit from more "[specialized] structures and processes for managing acquisitions" (Laamanen & Keil, 2008, p. 666) – two ways that cushion the effects of high activity load. Since "decentralization is equivalent to [an increase] . . . in the input of managerial services" (Penrose, 1959, pp. 49-50), scholars have hypothesized a firm's structure to yield effects similar to those of firm size, although empirical evidence on this has remained rather scarce. Second, another research stream (e.g., Castellaneta & Zollo, 2015; Laamanen & Keil, 2008) has emphasized the role of prior acquisition experience. Drawing on organizational learning theory, these studies have stressed that, through repetition of activities, firms build routines that allow them to reduce the attention managers *consciously* need to devote to these activities[16] (Ocasio, 1997) – a process that frees up managerial capacity (Castellaneta & Zollo, 2015; Laamanen & Keil, 2008). Third, a final set of studies (e.g., Zorn et al., 2019) has highlighted that the activity capacity available to acquirers also depends on the managerial capacity of the target firm. Specifically, Zorn et al. (2019) have presented evidence which shows that retaining target firm executives *after* the acquisition can alleviate the strains imposed by a high activity load, arguing that TMT retention increases the managerial capacity available for integration activities. Overall, these firm-level factors have strongly contributed to a more nuanced understanding of the effects of high levels of activity load than the concept of activity load alone.

As presented in Table 2, studies of activity load have extensively investigated the *performance* implications of information overload for acquirers, obtaining mostly consistent results. These studies have frequently stipulated and found a negative linear relationship between the activity load and

[14] I acknowledge that the level of interpretation difficulty can also depend on the "uncertainty, ambiguity, novelty, ... and intensity" of activities (Castellaneta & Zollo, 2015, p. 142). However, given that prior literature on activity load has mostly differentiated activities by their complexity (e.g., Castellaneta & Zollo, 2015; Collins et al., 2009; Ellis et al., 2011; Zorn et al., 2019), I follow this route to ensure comparability of results.

[15] Penrose (1959) argues that firms can only expand their activity capacity over time since "existing managerial personnel provide services that cannot be provided by personnel newly hired from outside the firm . . . because the experience they gain from working within the firm and with each other enables them to provide services that are uniquely valuable for the operations of [that firm]" (pp.41-42). Shaver (2006) agrees, stating that "the capacity effect is not necessarily binding in the long term because of the firm's ability to increase capacity" (p. 966).

[16] That is, managers switch from controlled processing to automatic processing of activities, saving cognitive capacity (Ocasio, 1997).

firm-level performance of acquirers, covering a wide array of research contexts, such as post-merger integration (PMI) (Barkema & Schijven, 2008; Shaver, 2006; Zorn et al., 2019) and acquisition programs (Kusewitt, 1985; Laamanen & Keil, 2008). Yet, a closer look at the literature reveals differences in the extent to which both activity load perspectives have been researched by scholars. For instance, extant literature has predominantly explored the effects of acquisition volume and complexity *in isolation* (e.g., Barkema & Schijven, 2008; Castellaneta & Zollo, 2015; Kusewitt, 1985; Vermeulen & Barkema, 2002; Zorn et al., 2019), mostly finding a negative linear relationship[17] between activity load and firm performance with moderately large effect sizes[18]. By contrast, studies which examine the *joint effect* of acquisition volume and complexity (e.g., Laamanen & Keil, 2008) have been scarce and only partially substantiated the existence of that effect[19]. Overall, despite these differences in perspective, extant research has demonstrated that activity load has important performance implications for acquirers.

However, Table 2 also clearly shows that extant literature has largely abstracted from firm-level consequences other than performance. For instance, only few studies (e.g., Barkema & Schijven, 2008) have examined that overload can also trigger a restructuring response of acquiring firms. Specifically, building on the behavioral theory of the firm (BTF) (Cyert & March, 1963; March & Simon, 1958; Simon, 1945), Barkema and Schijven (2008) have argued that acquirers initially seek solutions to integration problems within target firms (i.e., local search) but eventually need to cut complexity through restructuring (i.e., distant search) since the cumulative inefficiencies from past integrations create an overload situation. Their empirical results, which show that the number of acquisitions since an acquirer's last restructuring is positively related to that firm's likelihood to engage in renewed restructuring, support their logic. Likewise, evidence on the relationship between activity load and acquisition behavior of firms remains surprisingly scarce. That is, although one would intuitively expect high levels of activity load to affect the acquisition behavior of firms (e.g., stop in acquisitions due to indigestion - Kusewitt, 1985), no study has – to the best of my knowledge – hitherto investigated whether activity load influences acquisition behavior. This thesis therefore aims to complement prior activity load research, which is summarized in Figure 4, by studying whether activity load can help explain acquisition behavior of firms, in general, and deviations from temporal acquisition patterns, in particular.

2.4. Synthesis and Hypothesis Development

I derive three essential insights from reviewing the literature. First, acquisition pattern literature exhibits a gap in research on antecedents which explain *systematic deviations* from established acquisition patterns – despite a myriad of factors that are known to influence differences in the acquisitiveness of acquiring firms. Second, research on momentum converged towards a clear consensus on the repetitive momentum hypothesis, which, however, was both theoretically and methodologically challenged by recent studies (e.g., Beck et al., 2008). Third, studies of activity load have extensively investigated performance implications of high levels of activity load for acquirers, largely abstracting from alternative firm-level consequences, such as restructuring (e.g., Barkema & Schijven, 2008) and, in particular, changes in acquisition behavior.

By combining these insights, I argue that high levels of activity load cause acquiring firms to *systematically* deviate from their established acquisition patterns. If true, this relationship could explain acquisition behavior opposite to the repetitive momentum hypothesis. To validate this idea, I follow a three-step approach hereafter, with each step answering one of the subquestions shown earlier in Figure 1. First, I examine how the volume of acquisitions affects acquisition behavior – a relationship that solely draws on the logic of the computation perspective and, thus, serves as a good baseline effect of activity load. Second, I analyze how this baseline effect varies with the complexity of acquisitions, exploring the joint effect of volume (computation perspective) and complexity (interpretation perspective). Third, I investigate whether and how the absorptive capacity of acquirers, which I proxy in two different ways (i.e., acquirer size and organizational structure), alleviates the effect of activity load. By following these three steps, I can disentangle the effects of each activity load driver and develop a nuanced understanding of how activity load affects the acquisition behavior of firms. Figure 5 summarizes my approach.

2.4.1. Baseline Effect: Acquisition Volume

A change in acquisition behavior is the final step in a dynamic process in which firms face an overload situation caused by high levels of activity load. This process starts with the build-up of acquisition experience which acquirers gain from each acquisition they pursue (Haleblian & Finkelstein, 1999; Hayward, 2002), allowing them to successively build up the knowledge of and skills for managing acquisitions[20] (Laamanen & Keil, 2008). Over time, this accumula-

[17] Barkema and Shijven's (2008) study is an exception since it only examines activity load as a moderating variable, corroborating that a higher number of acquisitions strengthens the inverted u-shape relationship between the number of acquisitions since an acquirer's last restructuring and firm-level performance of the acquirer.

[18] For instance, Castellaneta and Zollo (2015) find that an activity load increase (i.e., increase in the number of private equity (PE) investments) by one standard deviation reduces an investment's IRR by ten percent (computation perspective). Likewise, Zorn et al. (2019) observe a USD 217m drop in the market value of acquirers with an asset book value of USD 1bn for each additional nested target within (i.e., higher complexity of) the focal target (interpretation perspective).

[19] Specifically, Laamanen and Keil (2008) find a significant effect for the product term of *acquisition rate* and program scope but not for the product term of *acquisition rate variability* and program scope.

[20] I assume that the time intervals between individual acquisitions are neither too short nor too long because too short (long) intervals make it harder (uneconomical) for firms to infer correct learnings from (codify) past acquisition experience (Hayward, 2002; Laamanen & Keil, 2008).

Table 2: Overview of Studies of Activity Load and Activity Capacity

Study	Literature Stream	Key Finding	Activity Load Perspective	Firm-Level Consequence
Penrose (1959)	Organizational Growth	The growth of a firm is determined by the managerial services available to that firm and those consumed by expansion activities.	-	-
Kusewitt (1985)	Acquisition Programs	A high *number* of acquisitions within a short time can lead to indigestion problems and is negatively associated with acquirer performance. Targets of *large relative size* and from *unrelated industries* are also negatively related to performance. A larger acquirer size and decentralized structure potentially alleviate this burden.	Computation and Interpretation[(a)]	Performance
Vermeulen and Barkema (2002)	International Expansion	A high/temporally concentrated *number* of new subsidiaries and a *broad product/ geographic* expansion scope create an overload that reduces a firm's absorptive capacity, weakening the positive relationship between international expansion and firm performance.	Computation and Interpretation[(a)]	Performance
Shaver (2006)	Post-Merger Integration	The limited cognitive capacity of managers can constrain the realization of synergies in the PMI phase.	-	Performance
Barkema and Schijven (2008)	Post-Merger Integration	The *number* of acquisitions since a firm's last restructuring is positively (negatively) related to further restructuring (performance). A higher acquisition intensity (more acquisition experience) strengthens (weakens) this effect.	Computation	Performance and Restructuring
Laamanen and Keil (2008)	Acquisition Programs	A high *volume*/uneven temporal distribution of acquisitions can overstrain the limited cognitive capacity of managers and inhibit the build-up of additional capacity, both being negatively related to acquirer performance. A *broader, more complex acquisition program scope* strengthens the negative effect of a high acquisition volume. A larger acquirer size and prior acquisition experience can weaken the negative effect of surges in acquisitions.	Computation and Interpretation[(b)]	Performance

(Continued)

Table 2—continued

Castellaneta and Zollo (2015)	Organizational Learning	A higher *number* of parallel PE investments (over-)saturates managers' limited attention capacity and is inversely related to PE investment returns. Prior acquisition experience can weaken this negative effect.	Computation	Performance
Zorn et al. (2019)	Post-Merger Integration	Nested acquisitions are *more complex* and, thus, require more managerial capacity than non-nested ones, reducing post-acquisition performance. Retaining TMT members of the focal target weakens this negative relationship.	Interpretation	Performance

Notes: Study investigates computation and interpretation perspective (a) seperately or (b) jointly and seperately.

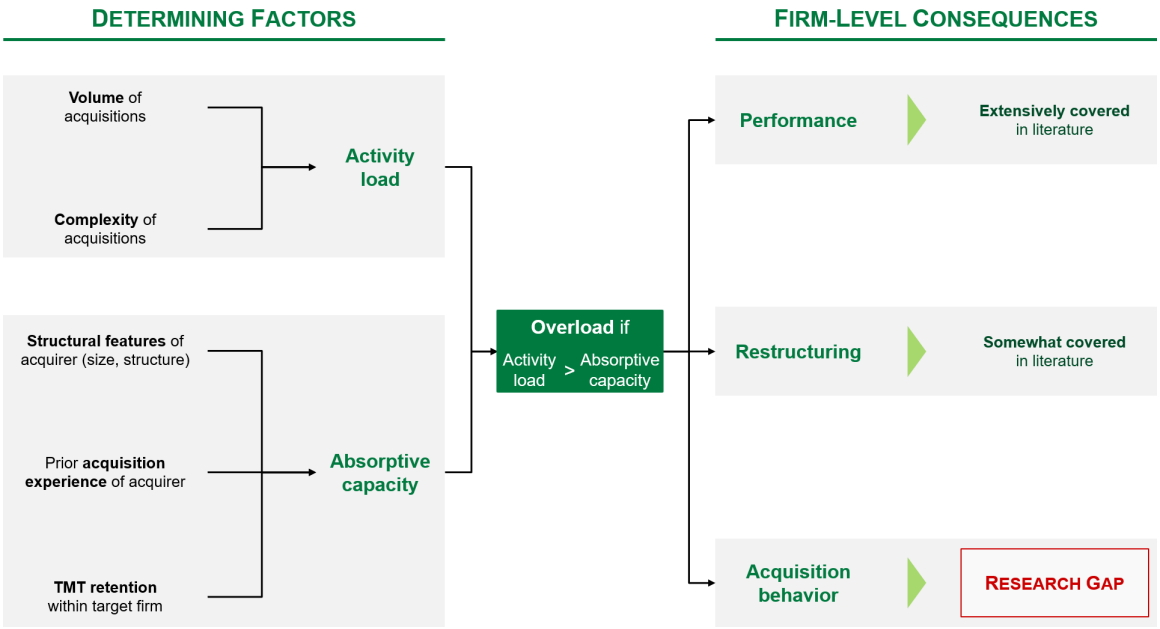


Figure 4: Determining Factors and Firm-Level Consequences of Overload

tion of acquisition experience leads to the formation of acquisition routines and a more central positioning of acquisitions in the cognitive maps of managers, creating repetitive momentum, which incentivizes firms to repeat learned behaviors and, thus, engage in further acquisitions (Amburgey & Miner, 1992). A higher number of subsequent acquisitions, however, directly translates into a higher activity load since acquirers not only need to integrate an ever-growing number of targets from past deals but also – in parallel – refill their deal pipeline by screening for and negotiating with potential targets to maintain acquisition momentum. In fact, the inertial pressures of momentum often are so pervasive (Amburgey & Miner, 1992; Miller & Friesen, 1980, 1982) that the steady increase in acquisition activity driven by momentum likely persists until the activity load from acquisitions exceeds the absorptive capacity of acquirers, creating a situation of information overload (Castellaneta & Zollo, 2015). As the strains of information overload are instantly felt by managers and, thus, must be immediately acted on, I expect the cognitive burden borne by managers in such a situation, *ceteris paribus*, to dominate the inertial pressures of momentum, forcing acquirers to reduce the activity load from acquisitions[21]. The simplest way for them to achieve this is by reducing their acquisition volume in the subsequent period – a change in acquisition behavior that would not only

[21] Expanding a firm's managerial capacity is not an option since it cannot be easily expanded in the short run (Penrose, 1959; Shaver, 2006).

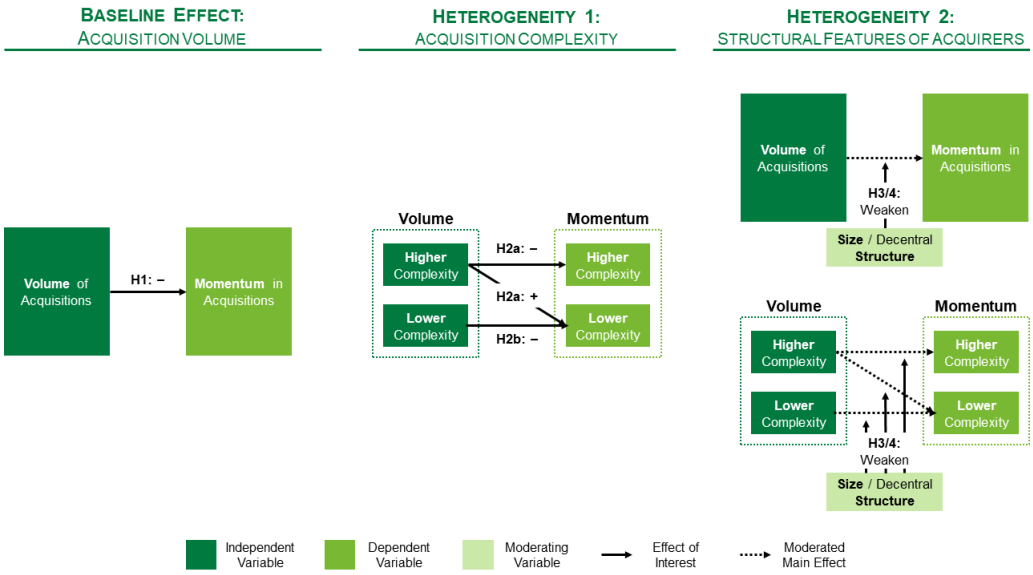


Figure 5: Key Relationships and Tested Hypotheses

allow acquirers to process their backlog of 'undigested' acquisitions (Kusewitt, 1985) but also correspond to a *discontinuation* in acquisition momentum (Beck et al., 2008; Vermeulen & Barkema, 2001). Stated formally:

> ***Hypothesis 1****: Increases in acquisition activity relative to a firm's past acquisition activity level reduce a firm's likelihood to engage in subsequent acquisitions.*

2.4.2. Heterogeneity in Acquirer Responses I: Acquisition Complexity

Because acquisitions are heterogeneous events (Haleblian & Finkelstein, 1999; Hayward, 2002) which differ in their rationale (e.g., Hayward, 2002), process management (e.g., Haspeslagh & Jemison, 1991), and performance (e.g., King et al., 2004), it is not surprising that some scholars (e.g., Zorn et al., 2019) have argued that acquisitions also vary in their degree of complexity and, thus, in their activity load. That is, depending on the characteristics of the target firm vis-à-vis the acquirer, acquisitions can vary in the amount of managerial capacity they consume (Zorn et al., 2019), allowing acquirers to exhibit cross-sectional variation in their behavioral response to an overload situation. Following this reasoning, acquisition research has widely discussed three dimensions of acquisition complexity, which operate on different levels *of aggregation and, thus,* capture different facets of acquisition complexity.

Figure 6 visualizes these dimensions, as well as the externally observable structural target firm attributes associated with them. That is, on the *country* level, acquisition complexity depends on the target's geographic location, which captures differences in cultural and institutional contexts between the home countries of the target and the acquirer (Collins et al., 2009; Ellis et al., 2011; Vermeulen & Barkema, 2002). Such differences add an additional layer of complexity to cross-border acquisitions since they pressure acquirers "to adapt home-grown mental maps, organizational structures, systems, and processes to the international setting" (Vermeulen & Barkema, 2002, p. 638) and absorb new "region-specific practices and business knowledge" (Ellis et al., 2011, p. 1265). Thus, cross-border acquisitions, on average, impose a higher activity load on acquirers than domestic ones. Likewise, on the *industry* level, acquisition complexity depends on the industry relatedness of the target and the acquirer. Since every industry operates on a different business logic (Prahalad & Bettis, 1986; Vermeulen & Barkema, 2002), which is reflected in different product offerings (Ellis et al., 2011) and internal structures (Finkelstein & Haleblian, 2002; Haleblian & Finkelstein, 1999), firms that acquire targets which operate in unrelated industries must manage "more complex . . . interdependencies across a wider variety of functions and products [during and after the acquisition]" (Ellis et al., 2011, p. 1264). Thus, unrelated acquisitions, on average, consume more managerial capacity than acquisitions of targets in related industries, imposing a higher activity load on acquiring firms (Zorn et al., 2019). Finally, on the *firm* level, acquisition complexity depends on the size of the target relative to the acquirer. That is, since the scale of operations of large targets requires large-scale integration activities, large acquisitions increase coordination costs for acquirers due to a higher number of "interrelated decisions . . . [and the] involvement of more . . . members across business[es]. . . and functional areas" (Ellis et al., 2011, p. 1263) and create a stronger disruptive effect within the acquirer's organization due to large-scale internal reorganization (Barkema & Schijven, 2008). Thus, large targets, on average, are more resource-consuming and complex to inte-

grate than smaller ones[22], imposing a higher activity load on firms (Zorn et al., 2019). Overall, the literary discourse on the dimensions of acquisition complexity highlights that the relationship between activity load and the acquisition behavior of acquirers can only be fully understood by examining the *joint* effect of acquisition volume and complexity (i.e., by combining both activity load perspectives).

Drawing on both activity load perspectives, I therefore argue that a firm's acquisition behavior response to information overload varies with the complexity level of acquisitions. Specifically, I expect firms to reduce the activity load from acquisitions in two ways. On the one hand, they can, *ceteris paribus*, cut the complexity of newly made acquisitions by switching from targets in a higher-complexity category (e.g., overseas targets) to ones in a lower-complexity category (e.g., domestic targets). By definition, such a response would represent a *deviation* from a firm's established acquisition pattern (e.g., shift from a pattern of overseas targets to one of domestic targets), allowing firms to maintain their overall acquisition momentum[23] while alleviating the strains of information overload. On the other hand, if acquirers have established an acquisition pattern of targets in a lower-complexity category, they have no choice to reduce their activity load other than by, *ceteris paribus*, decreasing the volume of newly made acquisitions. In doing so, firms can alleviate the strains of information overload only by discontinuing their acquisition momentum as there is no other target firm category to which they can switch. Thus, it follows:

> ***Hypothesis 2a**: Increases in acquisition activity within a higher-complexity target firm category relative to a firm's past acquisition activity level within that category reduce (increase) that firm's likelihood to engage in subsequent acquisitions within that higher-complexity (lower-complexity) category.*

> ***Hypothesis 2b**: Increases in acquisition activity within a lower-complexity target firm category relative to a firm's past acquisition activity level within that category reduce a firm's likelihood to engage in subsequent acquisitions within that category.*

2.4.3. Heterogeneity in Acquirer Responses II: Structural Features of Acquirers

Cross-sectional variation in the behavioral response of acquirers, however, could result not only from differences in acquisition complexity but also from differences in acquirer characteristics. Building on this idea, I argue that differences in acquirer size can explain why some acquirers can bear a higher activity load burden and, thus, are less likely to deviate from their stable acquisition pattern than others. That is, according to Penrosian resource-based logic, a larger firm size allows acquirers to access a larger pool of managerial resources and more specialized internal acquisition processes (Laamanen & Keil, 2008). Through this, acquirers *de facto* extend their absorptive capacity, which alleviates the strains imposed by high levels of activity load. This, in turn, reduces their likelihood of reaching a situation of information overload and, thus, helps them maintain their acquisition momentum. Following this rationale, I hypothesize:

> ***Hypothesis 3**: Increases in acquisition activity relative to a firm's past acquisition activity level less strongly affect that firm's likelihood to engage in subsequent acquisitions if that firm is larger.*

Using the same logic, I expect a more decentralized organizational structure to yield the same effect for acquirers as an increase in firm size or managerial capacity. That is, a higher degree of decentralization allows acquirers to distribute the activity load from acquisitions more evenly across organizational members, avoiding a too strong concentration of activity load within one single part of the organization. Like an increase in firm size, a more decentralized organizational structure expands the absorptive capacity of acquirers, alleviating the strains imposed by high levels of activity load. Consequently, I expect firms with a more decentralized structure, *ceteris paribus*, to be less likely to deviate from their established acquisition pattern. Thus, it follows:

> ***Hypothesis 4**: Increases in acquisition activity relative to a firm's past acquisition activity level less strongly affect that firm's likelihood to engage in subsequent acquisitions if that firm has a more decentralized organizational structure.*

3. Methodology

In the following, I outline the methodological approach taken to answer my research question. This outline consists of three parts. First, I delineate the data sources of my sample and the steps taken to modify the raw data. Second, definitions of all variables and the rationale for including them in

[22] I acknowledge that organizational learning literature (e.g., Castellaneta & Zollo, 2015; Ellis et al., 2011) has argued and found that repeated acquisitions of large targets (targets within the same geography/industry) reduce the *perceived* complexity of acquisitions of the same type due to routinization effects. However, I argue that routinization cannot eliminate the differences in complexity between larger (overseas/unrelated) and smaller (domestic/related) acquisitions for two reasons. First, routines are formed through repeated actions and, thus, need time to evolve (Collins et al., 2009). Second, if routines are not regularly used, past learnings which are not properly codified become irrelevant or are forgotten since employees, in which these learnings reside (Levitt & March, 1988), may move to different units or leave the company (Hayward, 2002).

[23] More specifically, by switching from targets in a higher-complexity category to targets in a lower one, firms shift their acquisition momentum from one target firm type to the other, discontinuing momentum for targets in the higher-complexity category while building up type-specific acquisition experience – and momentum – for targets in the lower-complexity category (Amburgey & Miner, 1992; Haleblian et al., 2006).

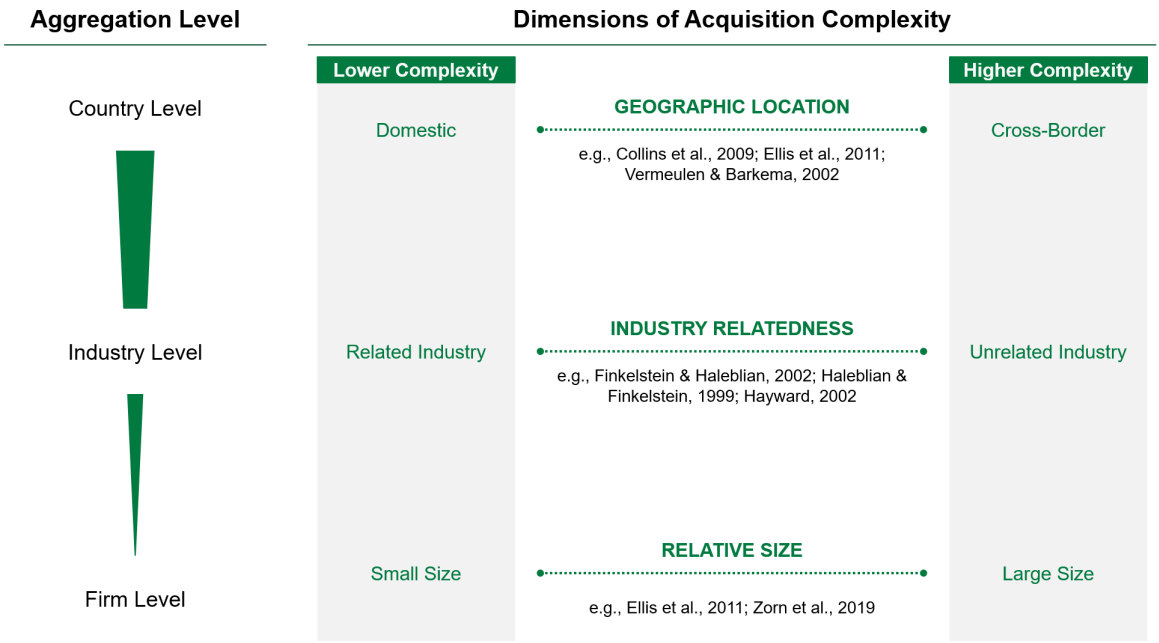


Figure 6: Dimensions of Acquisition Complexity by Aggregation Level

my analysis are presented.[24] Third, I elaborate on the empirical model used in this study.

3.1. Sample Sourcing and Data Cleaning

To corroborate my hypotheses, I compiled a panel dataset of publicly disclosed acquisitions of the 300 largest Fortune Global 500 firms over the 1990-2010 period. My rationale for choosing this particular sample is threefold. First, to ensure that all acquirers in the sample are sufficiently acquisitive, I only included acquirers that are large in size (Audia & Greve, 2006; Baysinger & Hoskisson, 1989). Firms listed in the Fortune Global 500 ranking meet this criterion particularly well since they are the largest ones by revenue and, thus, have sizable operations and acquisition activities. Second, I limited my sample to acquirers that were listed in the 1990 Fortune Global 500 ranking to avoid any survivorship bias. Third, to ensure that firms have acquisition streams that are sufficiently long for temporal dynamics to be visible, I selected a 21-year time horizon with high levels of historical acquisition activity (Cools et al., 2007; Kengelbach & Roos, 2011). Meeting all these criteria, my selected sample seems to be well suited to test my hypotheses.

This sample combines data from multiple data sources. For instance, in line with prior studies (e.g., Laamanen & Keil, 2008), all acquisition data stem from Refinitiv's Eikon M&A database (formerly Thomson One and Datastream), ensuring a comprehensive coverage of acquisitions over the sample period. In addition, to create the two controls *CEO overconfidence* and *CEO succession*, I extracted data from Compustat Execucomp. This dataset was complemented by human-coded data on TMT member titles as originally found on Execucomp, through which I proxied a firm's *degree of decentralization*. Finally, Standard & Poor's (S&P) Compustat North America database served as the source of company-level financials and business-segment data, which I used to create firm-level controls and cluster acquisitions by their observable attributes. All data on acquisitions, human-coded TMT member titles, and business-segment fundamentals were generously provided by Prof. Johannes Luger, who used them in prior research projects.

After merging these raw data into one comprehensive dataset, I executed several data cleans and transformations to avoid implausible data values and ensure consistency in my data. First, I eliminated duplicate M&A transactions in my acquisition raw data to ensure that deals in my final sample are unique, excluding 59 duplicate transactions. Second, I mapped acquisitions in a given calendar year (CY) to an acquirer's fiscal year (FY) with the same year identifier. That is, an acquisition occurring in August 2000, for instance, is counted towards an acquirer's FY 2000 even though the acquirer's fiscal year end (FYE) is in June, technically making that deal part of an acquirer's FY 2001. My rationale behind using this simplified mapping approach was twofold. On the one hand, only 45 (24 percent of) firms in my sample have a FYE that deviates from the calendar year end (CYE). On the other hand, since my research question investigates a phenomenon that evolves over multiple years, different FYEs likely do not distort my results. Third, I assumed that acquisitions with unreported deal values or target total assets were small and, thus, assigned them a value of zero for missing values of these variables. This simplifying assumption allowed me to include acquisitions with originally missing

[24] Please refer to Table A.1 in the appendix for detailed definitions and data sources of all variables used in this study.

data values that would have been dropped from my dataset otherwise, increasing my coverage. Fourth, I converted all firm-level financials reported in CAD to USD using the daily CAD/USD exchange rate at the respective reporting date provided by Compustat North America. This was necessary to ensure that all financials were denoted in the same currency and, thus, comparable.[25] Finally, I did not define control variables for firm-year observations with negative revenue or total asset values. This approach allowed me to only include observations with plausible values in my analyses while preserving the multi-year acquisition sequence of acquirers. In sum, these transformations and cleans resulted in a final baseline sample comprising 2,267 firm-year observations of 187 firms (i.e., an average of 12.1 years per firm), creating an unbalanced panel that is cross-section dominated (i.e., $N>T$), where N and T represent the number of firms and years, respectively.

3.2. Variable Definitions

3.2.1. Dependent Variables

To examine how the activity load from acquisitions affects the acquisition behavior of firms, I construct a binary measure of momentum which equals one if the sum of known deal values in period *t* is greater than or equal to the sum of known deal values in period *t-1* and zero otherwise. With this approach, I deviate from the methodology used in prior studies of momentum (e.g., Amburgey & Miner, 1992), which have used acquisition volume as a proxy for momentum. My rationale for this deviation is twofold. First, deal values likely are a more responsive indicator of (highly) short-term changes in acquisition behavior. To understand this, we need to consider a firm's option space during the negotiation phase of a deal. That is, if a firm intends to reduce its commitment to acquisitions (i.e., cut acquisition expenses), it can (i) adopt a more aggressive negotiation tactic to reduce the target's price or (ii) walk away from a deal. However, while the first option would still allow acquiring firms to close the deal[26], the second option would make it impossible for them to reach the strategic goal for which they initially pursued the deal. This suggests that firms are likely more flexible in adjusting the value of ongoing transactions than in changing their acquisition volume if they want to reach their *ex-ante* strategic goal with the currently negotiated deal. Second, historical patterns of acquisition volumes and deal values are fairly congruent[27] (Cools et al., 2007; Kengelbach & Roos, 2011), showing that increases (decreases) in acquisition volumes are often associated with increases (decreases) in deal values. In sum, both arguments indicate that deal values are a reasonable proxy for momentum.

Accounting for the three-level structure of my hypotheses[28], I use different operationalizations of momentum, depending on the tested hypothesis. Specifically, I distinguish between two sets of definitions. First, to test the baseline effect of acquisition volume (i.e., hypothesis 1), I use a firm's *total* known deal values as a measure of momentum, creating the binary variable *total acquisition momentum* which equals one if the sum of total known deal values in period *t* is greater than or equal to the sum of total known deal values in period *t-1* and zero otherwise. Second, to test how the baseline effect of acquisition volume varies with acquisition complexity (i.e., hypotheses 2a and 2b), I split a firm's *total acquisition momentum* by deal type (e.g., cross-border versus domestic), creating two mutually exclusive and collectively exhaustive (MECE)[29] subgroups of momentum – one subgroup of higher-complexity deals (i.e., *cross-border / cross-industry / large acquisition momentum*[30]) and another subgroup of lower-complexity deals (i.e., *domestic / within-own-industry / small acquisition momentum*). These deal-specific momentum measures are defined analogously to *total acquisition momentum* except that I use deal values of acquisitions of the respective deal type in the construction of these measures (e.g., deal values of cross-border deals for *cross-border acquisition momentum*). Finally, to test how the prior two effects vary with structural features of acquirers (i.e., hypotheses 3 and 4), I run models with both sets of momentum definitions (i.e., *total acquisition momentum* and deal-specific measures).

3.2.2. Independent Variables

Building on prior research (e.g., Castellaneta & Zollo, 2015; Laamanen & Keil, 2008; Zorn et al., 2019), I proxy the concept of activity load with a firm's number of completed acquisitions in the period *t-2* to *t*, choosing a three-year time window to account for a time lag between a deal's closing date and the completion of integration activities. As I did with my dependent variables, I distinguish between two sets of definitions. That is, on the one hand, I measure the activity load of a firm as that firm's *total acquisition activity* (i.e., its *total* number of completed acquisitions in years *t-2* to *t*), allowing me to test hypothesis 1, which examines the baseline effect of acquisition volume on acquisition behavior. On the other hand, to test hypotheses 2a and 2b, I split a firm's *total acquisition activity* by deal type, creating two MECE subgroups of acquisitions which differ by their relative complex-

[25] Specifically, Compustat displays financial data in the company's reported currency, whereas Refinitiv Eikon reports acquisition data only in USD. This is a problem when calculating the target-asset-to-acquirer-asset ratio as some acquirers only report their financials in CAD.

[26] Provided the target does not walk away from the deal.

[27] See Figure A.1 and Figure A.2 in the appendix.

[28] That is, the baseline effect of acquisition volume (first level), heterogeneity in that baseline effect due to differences in acquisition complexity (second level), and heterogeneity in the prior two effects due to differences in structural features of acquiring firms (third level).

[29] Please note that momentum measures *within* a subgroup are *not* MECE. For instance, a deal can be a cross-border deal and a large deal at the same time. This overlap is unproblematic because I only regress deal-specific momentum measures on acquisition activity measures of the same complexity dimension (e.g., *cross-border acquisition momentum* on *cross-border acquisition activity*).

[30] Please refer to Table A.1 in the appendix for precise definitions of the terms 'cross-border', 'cross-industry', and 'large'.

ity level (i.e., higher-complexity deals, which include *cross-border (cross-industry; large) acquisition activity*, and lower-complexity deals, which cover *domestic (within-own-industry; small) acquisition activity*) and allow me to assess the *joint* effect of acquisition volume and complexity. Finally, to ensure consistency with my dependent variable definitions, I run models with both sets of activity load definitions (i.e., *total acquisition activity* and deal-specific acquisition activity measures) to test hypotheses 3 and 4.

3.2.3. Control Variables

To account for alternative explanations of acquisition behavior, I include an extensive set of control variables which captures the effect of specific acquisition characteristics, acquiring firm characteristics, and industry-level acquisition activity on the acquisition behavior of acquiring firms. For this, I followed a two-step selection procedure to ensure only relevant variables are incorporated in the model. First, I screened extant literature for factors that are known to affect a firm's acquisition behavior and selected those for which past studies provided solid theoretical arguments and, ideally, robust empirical evidence.[31] With this, I directly responded to King et al. (2004) who called for greater consistency of empirical measurement in acquisition research, ensuring comparability of results across studies. Second, in case prior acquisition literature did not cover factors which are core to my hypothesis development, such as *proportion of large acquisitions* and *acquirer degree of decentralization*, I included such factors as novel variables. The following control variables are the result of this two-step selection procedure:

Acquisition complexity. Scholars who adopt the interpretation perspective of activity load (e.g., Zorn et al., 2019) have found that the acquisition behavior of firms can vary with the complexity level of acquisitions. To account for this effect, I include three distinct proxies of acquisition complexity.[32] First, in line with prior literature (e.g., Laamanen & Keil, 2008), I control for the *proportion of cross-border acquisitions* made by firm *i* in years *t-2* to *t*, where the term 'cross-border' refers to acquisitions of targets *not* based in the US. This variable captures differences in cultural and institutional contexts of targets and acquirers that drive acquisition complexity and, thus, a firm's activity load from acquisitions. Second, I include the *proportion of cross-industry acquisitions* over the same period and expect cross-industry acquisitions to be more complex than acquisitions within the acquirer's own industry due to differences in business logic and organizational setups across industries (e.g., Prahalad & Bettis, 1986; Vermeulen & Barkema, 2002). Following prior studies (e.g., Ellis et al., 2011; Laamanen & Keil, 2008; Zorn et al., 2019), I classify acquisitions as 'cross-industry acquisitions' if the first two digits of the target's and acquirer's primary SIC codes are not identical. Third, I control for the *proportion of large acquisitions* from *t-2* to *t* since the integration of larger targets imposes a higher strain on the acquirer's organization than the integration of smaller ones (Barkema & Schijven, 2008; Ellis et al., 2011) and, thus, limits the availability of organizational resources for subsequent acquisitions. Targets are considered 'large' if their relative size (measured relative to acquirer total assets) is greater than or equal to the sample mean of 8.9 percent.

CEO overconfidence. Scholars have found that overconfident CEOs overestimate their abilities and, thus, are more likely to engage in acquisitions than non-overconfident CEOs (Malmendier & Tate, 2005, 2008). I account for this finding by adding *CEO overconfidence* to the model, using the stock-option-based measure developed by Malmendier and Tate (2005, 2008), which assumes that late exercise of stock options signals higher confidence. Specifically, I define *CEO overconfidence* as a dummy variable which is equal to one if a CEO has not exercised at least 67 percent of his/her exercisable in-the-money stock options in a given year and zero otherwise.

CEO succession. A phenomenon extensively discussed by management scholars (e.g., Karaevli, 2007; Zajac, 1990), *CEO succession* not only affects the performance of firms but also likely influences their acquisition behavior. This is because CEOs, as individuals, differ in their level of risk appetite and, thus, likely show variation in the strategic decisions they make, including acquisitions. To account for this effect, I include a dummy variable for *CEO succession* which is equal to one if a firm experienced a CEO change in a given year and zero otherwise.

Degree of decentralization. I control for a firm's *degree of decentralization* to account for the Penrosian argument that a more decentralized organizational structure allows firms to distribute the workload from acquisitions across a larger pool of managerial resources and, thus, absorb the activity load from acquisitions more effectively. For this, I use the percentage of executives with divisional or geographic titles (as opposed to functional, matrix, or general manager titles) in period *t*, where a value of one (zero) represents a fully decentralized (centralized) structure.

Performance relative to aspirations. Prior acquisition research has shown that the performance of firms relative to their own prior performance (i.e., historical aspiration level) affects their acquisition activity such that acquisition activity increases (falls) if firms are performing below (above) their historical aspiration level (Iyer & Miller, 2008). In line with past studies, I measure a firm's performance relative to its historical aspiration level via the recursive formula $P_t - A_t$, where P_t is a firm's ROA in period t and A_t is a firm's aspiration level in period t that is given by $0.3\ P_{t-1} + 0.7\ A_{t-1}$ (Greve, 2002), and model it as a spline function (i.e., two distinct variables (i) *Performance above aspirations* and (ii)

[31] See section 2.1 for a review. Please note that I only included variables for which I had database access.

[32] I only include measures of acquisition complexity that are *not* captured by the acquisition activity measure in the given model. For instance, I include *proportion of cross-industry acquisitions* and *proportion of large acquisitions* in models in which I regress *cross-border acquisition momentum* on *cross-border acquisition activity*, omitting *proportion of cross-border acquisitions*, which is already captured by *cross-border acquisition activity*.

Performance below aspirations) (Shinkle, 2012).[33]

Growth desperation. Kim et al. (2011) have presented evidence that firms face higher pressure to engage in acquisitions (and overpay) if their organic growth rate is below that of prior years and if they have historically grown stronger through acquisitions than their industry peers. To account for the effect of *growth desperation* on a firm's acquisition behavior, I incorporate the growth of firms relative to their own prior growth as a control, measuring it as a spline function analogously to *performance above/below aspirations* (i.e., (i) *growth above aspirations* and (ii) *growth below aspirations*) but using a firm's year-on-year sales growth instead of its ROA.

Slack resources. According to the BTF (e.g., Cyert & March, 1963; Levinthal & March, 1981), excess resources give firms more room to experiment and pursue new strategic opportunities, such as acquisitions. Recent empirical work has confirmed this by showing that resource slack is positively related to a firm's acquisition propensity (Iyer & Miller, 2008). Yet, contrary to prior studies, I refrain from using a firm's debt-to-equity (D/E) ratio and instead take a firm's debt-to-total-assets ratio as an inverse proxy for *slack resources* since the total equity of a substantial proportion (∼7 percent) of firm-years in my sample is negative. This allows me to keep firm-years with negative equity in my sample and avoid implausible, negative D/E values. Correspondingly, a higher debt-to-total-assets ratio signals a lower level of slack resources.

Acquirer size. Since size is related to the number of resources available to firms (e.g., Laamanen & Keil, 2008; Penrose, 1959) and their risk appetite (e.g., Audia & Greve, 2006; Baysinger & Hoskisson, 1989), I control for firm size by including a proxy which is often used in literature: The natural logarithm of total assets (e.g., Iyer & Miller, 2008; Laamanen & Keil, 2008).

Diversification. Diversified firms may be exposed to more acquisition opportunities than their undiversified peers (Sanders, 2001). I account for this by including the control *diversification*, which I operationalize with the Jacquemin-Berry entropy index (Palepu, 1985; Sanders, 2001).

Industry acquisition activity. The acquisition activity within an industry can strongly affect an individual firm's acquisition behavior (e.g., Baum et al., 2000; McNamara et al., 2008) due to imitation/bandwagon effects and exhaustion of potential targets. To ensure these factors do not drive my results, I control for *industry acquisition activity* by including the average number of acquisitions per industry (based on first two digits of primary SIC code) in the period *t-2* to *t*.

Year dummies. To account for potential contemporaneous correlation (i.e., residuals of firms *i* and *j* are correlated in period *t*), I include a full set of year dummies in all model specifications (excluding one year for identification purposes) (Certo & Semadeni, 2006). Contemporaneous correlation can severely bias estimates in panel regression and is of particular concern in cross-section-dominated panel data like mine (Certo & Semadeni, 2006; Wooldridge, 2010).

3.3. Estimation Strategy

To estimate the effect of activity load from acquisitions on acquisition behavior, I constructed the following non-linear regression model which uses the logistic distribution as a link function and the maximum likelihood (ML) method for parameter estimation (Wooldridge, 2010):

$$P(Momentum_{i,t} = 1 \,|\, A_{i,t-1}, \Gamma_{i,t-1}, T_t) = \Lambda(\beta_0 + \beta_1 A_{i,t-1} + \delta_1 \Gamma_{i,t-1} + \delta_2 T_t + \varepsilon_{i,t}),$$

where $\Lambda(x) = e^x/(1+e^x)$ is the logistic function, A is a vector of acquisition variables, Γ is a vector of time-varying controls, and T is a full set of year dummies (excluding the base year). Subscripts t and i denote the fiscal year in which the acquisition was completed and the acquirer, respectively. In addition, all independent variables (except year dummies) are lagged by one period to eliminate concerns about endogeneity[34] (Dobbins & Jacob, 2016; Iyer & Miller, 2008).

Albeit the use of a binary dependent variable does not immediately call for a non-linear model specification (Wooldridge, 2010), I preferred a non-linear model over a linear one, such as a linear probability model (LPM). My rationale for this was twofold. First, given the nature of my data, a non-linear model produces better estimates of the marginal effects of my covariates than an LPM. That is, an LPM is suitable "if most . . . [covariates] are discrete and take on only a few values" (Wooldridge, 2010, p. 564). Yet, many of my covariates (e.g., activity load) have continuous – and sometimes even extreme – values that violate this assumption.[35] This conclusion is supported by the fact that roughly a third of my baseline sample observations have predicted probabilities that lie outside the unit interval if an LPM is used (Wooldridge, 2010). Second, consistent with prior studies of momentum (e.g., Beck et al., 2008), I used a logit model to ensure comparability of results across studies. Thus, a non-linear model seems warranted.

Building on this specification, I selected a hybrid logit model[36] – an estimation strategy that has received increased attention by management scholars (e.g., Allison, 2005, 2009;

[33] Parameter a=0.3 used to ensure consistency with prior literature (e.g., Iyer & Miller, 2008). Iyer and Miller (2008) chose a=0.3 because this provided the best model fit. Also, like Kim et al. (2015), I set a firm's historical aspirations to zero the first time a firm entered the sample.

[34] This approach follows the one of Iyer and Miller (2008), who model the effect of performance feedback on the likelihood of an acquisition. This is necessary to satisfy a key condition for causality. That is, if a relationship XY is to be causal, a change in X must precede a change in Y (Wolfolds & Siegel, 2019). Given that I define momentum (Y) as a dummy which indicates the change in deal values from *t-1* to the focal year *t*, I can only include acquisitions (X) that occur *before* the focal year *t* (i.e., *t-3* to *t-1*). Likewise, controls based on *end-of-year* balance sheet data, such as ln(total assets), cannot explain a change in momentum occurring *during* period *t*. Consequently, I lag these variables, too.

[35] Please refer to Table A.2 in the appendix for a detailed set of descriptive statistics for all variables used in the baseline model.

[36] Please note that I performed all my analyses with a manually constructed

Certo et al., 2017; Schunck, 2013; Schunck & Perales, 2017). A hybrid model allows for the simultaneous estimation of within-unit effects (i.e., changes that occur *within* units over time) and between-unit effects (i.e., changes that occur *between* units) within the same model by decomposing each independent variable into two distinct variables: (i) a group-centered variable (i.e., within-unit effect) and (ii) a group mean variable (i.e., between-unit effect) (Certo et al., 2017; Schunck, 2013; Schunck & Perales, 2017). A random effects model is then used to estimate the within- and between-unit effects for each independent variable (Certo et al., 2017). Correspondingly, to disentangle these two effects, vectors A and Γ include both a firm-centered variable and a variable for the time-invariant firm mean for each independent variable as distinct regressors.

At first glance, a hybrid logit model seems unusual since my research question solely explores a within-firm relationship. Such a relationship can be easily analyzed with a fixed effects model, which would eliminate unobserved time-invariant heterogeneity at the firm level (Dobbins & Jacob, 2016; Herold, 2019; Wooldridge, 2010). Also, a fixed effects logit model would ensure methodological consistency with prior acquisition literature, following the call of Beck et al. (2008), who have shown that prior literature on momentum has likely produced biased results due to overreliance on random effects models. A fixed effects model therefore seems appealing.

Yet, two reasons support the hybrid model specification. First, from a theoretical standpoint, the parameter estimates of a hybrid logit model are approximately identical to those of a fixed effects logit model (Allison, 2009; Schunck & Perales, 2017).[37] The difference is the estimation technique. That is, whereas the hybrid logit model uses a parametric approach to estimate the fixed effects (i.e., includes them as separate regressors in the model), the fixed effects logit model uses a more restrictive, non-parametric approach (i.e., a conditional ML estimator) to eliminate unobserved time-invariant heterogeneity at the firm level (Allison, 2009; Schunck & Perales, 2017; Wooldridge, 2010). Second, from a practical standpoint, a hybrid logit model allows for the estimation of partial effects on response probabilities (Wooldridge, 2010). These cannot be readily estimated in a fixed effects logit model since doing so would require either a precise value or a distribution for the fixed effect to be specified (Wooldridge, 2010). However, since momentum theory does not justify any exact value or distribution for the fixed effect, neither of these options are plausible if I aim to quantify the partial effects on my response probabilities. Due to these points, I favored a hybrid logit model over a fixed effects alternative.

hybrid logit model, which I preferred over STATA's *-xthybrid-* command for two reasons: (i) *-xthybrid-* allows for neither the *-margins-* command nor any factor notation of variables (Schunck, 2013), which are both critical for estimating interactions, and (ii) the results of the manually constructed model are identical to those of the *-xthybrid-* model.

[37] For comparison, results of a fixed effects logit model are shown in Table A.4 in the appendix.

4. Empirical Results

In this section, I present the results of my empirical analysis in three steps. First, I explore the summary statistics and correlations of my raw data to find first indications for the existence of my postulated relationships. Second, I dive into my main regression analysis. Finally, in my robustness tests, I test for the robustness of the relationships discovered in my main analysis.

4.1. Summary Statistics

Table 3 presents descriptive statistics and pairwise correlations of all variables used in my main analysis. All statistics are based on variables in their raw form (i.e., not split into their within- and between-firm components) and, thus, reflect both within- and between-firm variance.

The pairwise correlations in Table 3 only offer ambiguous initial support for the existence of my hypothesized relationships. Two observations lead to this conclusion. First, while bivariate correlations between momentum and acquisition activity variables within the same target firm category are significant ($p<.01$), these variables are *positively* correlated (colored in light green in Table 3). These relationships thus have a sign opposite to the one predicted in hypotheses 1 and 2, tentatively indicating the existence of repetitive momentum (Amburgey & Miner, 1992). However, since the correlation coefficients in Table 3 include both within- and between-firm variance, the overall positive correlations could stem from a positive between-firm effect (i.e., repetitive momentum) offsetting a negative within-firm effect (i.e., momentum discontinuation due to high levels of activity load).[38] Thus, the correlation matrix shown in Table 3 does not provide sufficiently fine-grained information to indicate the existence of my postulated within-firm relationships. Second, the correlations between momentum in a lower-complexity target firm category (e.g., *domestic acquisition momentum*) and acquisitions in a higher-complexity target firm category (e.g., *cross-border acquisition activity*) are positive and mostly significant ($p<.01$) (colored in dark green in Table 3). Although this can indicate the existence of a firm's 'switching response' as predicted in hypothesis 2a, these correlations include both within- and between-firm variance, making their interpretation ambiguous in my research setting. In sum, the correlations in Table 3 do not clearly indicate the existence of my postulated relationships.

Table 3 further reveals that multicollinearity issues are unlikely to arise in my analysis as most correlations are either rather weak (i.e., below $|.30|$) or moderate (i.e., around $|.50|$) (Judge et al., 1982). This holds true even though strong pairwise correlations between momentum (acquisition activity) variables, with values of 0.83 (1.00), would initially suggest the opposite. In fact, these high values are not

[38] In fact, the correlations between momentum and acquisition activity in Table 3 should be positive. This is because prior momentum studies have found a positive relationship between acquisition activity and momentum when using random effects models (Amburgey & Miner, 1992).

Table 3: Descriptive Statistics and Correlation Matrix

Variable	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1. Total Acquisition Momentum	0.38	0.48																											
2. Cross-Border Acquisition Momentum	0.25	0.43	0.43***																										
3. Domestic Acquisition Momentum	0.32	0.47	0.77***	0.11***																									
4. Cross-Industry Acquisition Momentum	030	0.46	0.60***	0.35***	0.54***																								
S. Within-Own-Industry Acquisition Momentum	0.26	0.44	0.61***	0.36***	0.50***	0.09***																							
6. Large Acquisition Momentum	0.11	0.31	0.38***	0.08***	0.41***	0.20***	0.33***																						
7. Small Acquisition Momentum	0.35	0.48	0.83***	0.46***	0.61***	0.62***	0.49***	0.06***																					
8. Total Acquisition Activity	14.15	23.35	0.06***	0.20***	0.11***	0.13***	0.05**	0.06***	0.09***																				
9. Cross-Border Acquisition Activity	6.29	12.26	0.05**	0.18***	0.09***	0.11***	0.04*	0.05**	0.08***	0.95***																			
10. Domestic Acquisition Activity	7.86	12.37	0.06***	0.19***	0.12***	0.14***	0.05**	0.07***	0.09***	0.95***	0.80***																		
11. Cross-Industry Acquisition Activity	9.53	21.25	0.05**	0.16***	0.09***	0.11***	0.00	0.04*	0.07***	0.95***	0.89***	0.90***																	
12. Within-Own-Industry Acquisition Activity	4.62	7.59	0.04**	0.15***	0.10***	0.09***	0.15***	0.09***	0.06***	0.43***	0.42***	0.39***	0.11***																
13. Large Acquisition Activity	0.53	0.92	-0.01	0.04**	0.02	0.07***	0.03	0.14***	0.07***	0.18***	0.13***	0.22***	0.11***	0.25***															
14. small Acquisition Activity.	13.63	23.20	0.06***	0.20***	0.11***	0.13***	0.05**	0.06***	0.09***	1.00***	0.95***	0.95***	0.95***	0.42***	0.14***														
15. Proportion of Cross-Border Acquisitions	0.42	0.30	-0.01	0.03	-0.05**	-0.02	-0.02	00	-0.02	0.04**	0.21***	-0.13***	02	0.08***	-0.11***	0.05**													
16. Proportion of Cross-Mum Acquisitions	0.60	0.32	0.04**	0.06***	0.06***	0.12***	-0.06***	-0.01	0.04**	0.14***	0.10***	0.17***	0.27***	-0.31***	-0.06***	0.15***	-0.05***												
17 Proportion of Large Acquisitions	0.08	0.18	-0.08***	-0.09***	-0.08***	-0.08***	-0.05**	02	-04**	-0.14***	-0.13***	-0.12***	-0.12***	-0.09***	0.47***	-0.15***	-0.20***	-0.20***											
18 CEO Overconfidence	0.68	0.47	0.04**	0.06***	0.04**	0.05**	0.04*	00	0.05**	0.06***	0.07***	0.04*	0.05**	0.05**	01	0.06***	0.03	-0.01	-0.02										
19. CEO Succession	0.13	0.34	0.00	0.03	0.00	0.00	0.02	0.00	0.01	0.00	0.01	-0.01	0.00	0.01	-0.02	0.00	0.03	0.01	-0.04*	-0.04*									
20. Degree of Decentralization	0.19	0.18	0.00	-0.02	-0.01	-0.01	01	02	-0.01	-0.04**	-0.06***	-0.03	-0.04*	-0.04*	0.00	-0.04**	-0.03	-0.03	-0.01	-0.01	0.00								
21 Performance Above Aspirations	0.02	0.03	0.08***	0.01	0.09***	0.09***	0.05**	0.06***	0.07***	-0.10***	-0.09***	-0.10***	-0.10***	-0.04**	0.04**	-0.10***	-0.05***	-0.08***	0.13***	0.05**	-0.03	-0.05**							
22. Performance Below Aspirations	-0.01	0.03	0.11***	0.08***	0.10***	0.08***	0.06***	0.06***	0.09***	02	03	02	03	-01	-0.03	02	02	03	-0.02	0.10***	-02	-0.09***	0.33***						
23. Growth Above Aspirations	0.06	0.11	-0.06***	-0.04**	-0.03	-0.01	-0.03	-0.03	-0.03	-0.02	-0.04**	-0.01	-0.02	-0.02	0.13***	-0.03	-0.07***	-0.01	0.12***	0.04**	0.00	0.02	0.14***	0.00					
24. Growth Below Aspirations	-0.06	0.13	0.05**	0.07***	0.04*	03	0.06***	-0.01	0.06***	01	02	00	00	03	01	01	01	-0.04*	00	0.04**	-03	-0.03*	0.17***	0.35***	0.25***				
25. Slack Resources	0.64	0.17	-0.09***	-0.05**	-0.12***	-0.11***	-0.07***	-0.07***	-0.08***	01	0.05**	-0.02	0.04*	-0.06***	-0.06***	02	0.06***	-0.03	0.04**	-0.02	0.04*	0.00	-0.14***	-0.04**	-0.05**	0.03			
26. Acquirer Size	9.00	1.28	0.08***	0.22***	0.09***	0.13***	0.10***	0.04*	0.12***	0.55***	0.53***	0.51***	0.48***	0.35***	0.13***	0.55***	0.11***	0.15***	-0.18***	0.07***	00	-0.01	-0.13***	-0.01	-0.04*	-0.05**	0.02		
27. Diversification	0.90	0.48	0.00	0.05**	0.01	0.05***	-0.03	-0.02	02	0.27***	0.22***	0.29***	0.27***	0.06***	01	0.27***	-0.04*	0.22***	-0.13***	01	00	01	-0.17***	0.03*	-0.05**	01	0.04*	0.25***	
28. industry Acquisition Activity	44.97	35.71	0.05**	0.13***	0.08***	0.06***	0.13***	0.07***	0.08***	0.30***	0.28***	0.30***	0.24***	0.26***	0.18***	0.30***	0.03*	-0.03	-0.08***	0.08***	0.05'	0.05***	-0.09***	-0.01	-0.01	0.00	0.00	0.24***	0.13***

Notes: N=2,267. All statistics are based on raw variables (i.e., before decomposing independent variables into their within- and between-firm components). Light green cells indicate correlations between momentum and acquisition activity variables within the same target firm category (i.e, hypotheses 1 and 2). Conversely, dark green cells indicate correlations between momentum in a lower-complexity target firm category and acquisition activity in a higher-complexity target firm category (i.e., second part of hypothesis 2a).
$^{***}p < 0.01$, $^{**}p < 0.05$; $^{*}p < 0.1$.

surprising as all momentum (acquisition activity) variables which proxy different acquisition complexity dimensions are subcomponents of *total acquisition momentum* (*total acquisition activity*). As such, these variables are strongly correlated with each other by definition. In addition, 91.1 percent of acquisitions in my baseline sample are small, thus explaining the particularly strong correlations between *small acquisition momentum* (*small acquisition activity*) and all other momentum (acquisition activity) variables. Furthermore, I only include one (two MECE) momentum (acquisition activity) variable(s) in the same model, creating a setup in which multicollinearity issues are unlikely to arise.

To corroborate the absence of multicollinearity, I ran variance inflation factor (VIF) tests of all variables which are split into their within- and between-firm components and used in my main model (see section 4.2). Table A.3 in the appendix shows that (almost) all variables have a VIF statistic that is below the advocated threshold value of (five) ten (Neter et al., 1989). The only two variables with VIF coefficients of approximately five are the between-firm components of *cross-border acquisition activity* and *domestic acquisition activity*. However, I decided to retain these in the model for consistency with all other model specifications. From these results, I thus conclude that multicollinearity issues are unlikely to arise in my analysis.

4.2. Main Analysis

Table 4 presents the empirical results of my main analysis. For simplicity, only the coefficients for the group-centered variables (i.e., within-firm effects) are reported as my theoretical interest solely lies in those. Models 1 and 2 in Table 4 use an acquirer's *total acquisition momentum* as the dependent variable to investigate the baseline effect of activity load. Models 3 to 14 go one step further and address six different momentum definitions, which are based on observable target firm attributes, to measure heterogeneity in acquirer responses due to differences in acquisition complexity. Furthermore, all odd-numbered (even-numbered) model specifications include controls only (all independent variables), enabling sanity checks of the effect directions and effect sizes of controls across models with the same dependent variable (e.g., to detect common-factor multicollinearity (Kalnins, 2018)). Finally, in line with reporting standards for logit models, all coefficients represent odds ratios – exponentiated coefficients that express the e^{β}-times change in the odds of an event due to a one-unit change in a variable (Hoetker, 2007).[39]

Recalling hypothesis 1, I expect an increase in an acquirer's acquisition activity relative to that firm's past acquisition activity level to lead to a discontinuation in acquisition momentum. That is, the greater the increase in the number of acquisitions in a given acquisition stream, the higher the activity load borne by the acquiring firm's managers. This increase in activity load, in turn, increases the likelihood of an information overload, which eventually forces acquirers to reduce the number of acquisitions to alleviate the strains associated with this overload. The significant ($p<.05$) and smaller-than-one odds ratio in Model 2 (highlighted in light green in Table 4) shows a decrease in the odds of *total acquisition momentum*, thereby confirming hypothesis 1.

Odds ratios, however, are not very informative if I seek to estimate the effect of a variable on the *predicted probability* of a given event (Hoetker, 2007) – "the natural metric of the dependent variable [in a logit model]" (Mize, 2019, p. 84). To do this, I estimate the marginal effects of my variables of interest, using two best practice approaches advocated in recent methodological studies (e.g., Hoetker, 2007; Mize, 2019; Mize et al., 2019). First, following Hoetker (2007) and Train (1986, 2009), I calculate the average marginal effects (AME) of my variables of interest in Table 5 to estimate the average response across my sample. That is, I compute the marginal effects for every firm-year and average these across my sample (Mize, 2019), finding that a one-unit increase in the number of acquisitions relative to a firm's past acquisition level, on average, is associated with a 0.3 percentage-point (pp) decrease in the predicted probability of that firm's *total acquisition momentum*. Second, to account for the fact that the effect of a change in any variable varies with the values of all covariates (i.e., with the initial likelihood of an event) (Hoetker, 2007), I plot the relationship between a change in a firm's *total acquisition activity*[40] (i.e., within-firm effect) and that firm's probability of *total acquisition momentum* in Figure 7 (e.g., Mize, 2019). The downward-sloping area in Figure 7 corroborates that larger increases in acquisition activity are associated with lower likelihoods of acquisition momentum. In sum, these two best practices allow me to properly estimate and interpret the effect of activity load on the probability of acquisition momentum, ensuring a solid understanding of my results.

In hypothesis 2a, I predict a negative (positive) relationship between an acquirer's activity load in a higher-complexity target firm category and that firm's acquisition momentum within that category (in a lower-complexity category). Specifically, since acquisitions vary in their relative complexity level, acquirers can, *ceteris paribus*, cut the activity load of newly made acquisitions by switching from targets in a higher-complexity category (e.g., overseas targets) to those in a lower-complexity category (e.g., domestic targets). This would equal a deviation from a firm's established acquisition pattern (e.g., shift from a pattern of overseas targets to one of domestic targets) that allows acquirers to maintain their overall acquisition momentum while alleviating the strains of information overload. However, my results in Table 4 only provide mixed support for this prediction. That is, while I find significant ($p<.01$) and smaller-than-one odds ratios in Model 4 and Model 12, the smaller-than-one

[39] In line with the recommendations presented by Hoetker (2007), I omitted measures of model fit from Table 4 since all available pseudo-R^2 measures for logistic regressions do not equal the R^2 used in ordinary least squares (OLS) regression, thus avoiding potential sources of confusion.

[40] To eliminate the effect of extreme outliers, I limit the plot to values between the 1st percentile and 99th percentile of my independent variable.

Table 4: Hybrid Panel Logit Regression Analysis with Different Acquisition Momentum Definitions as the Dependent Variable

	Total Acquisition Momentum		Cross-Border Acquisition Momentum		Domestic Acquisition Momentum		Cross-Industry Acquisition Momentum		Within-Own--Industry Acquisition		Large Acquisition Momentum		Small Acquisition Momentum	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12	Model 13	Model 14
Hypotheses														
Total Acquisition Activity		0.987**												
		(0.005)												
Cross-Border Acquisition Activity				0.962***		1.017								
				(0.011)		(0.011)								
Domestic Acquisition Activity				1.023**		0.979**								
				(0.010)		(0.009)								
Cross-Industry Acquisition Activity								0.994		1.007				
								(0.006)		(0.009)				
Within-Own-Industry Acquisition Activity								1.013		0.931**				
								(0.017)		(0.017)				
Large Acquisition Activity												0.747***		1.062
												(0.066)		(0.070)
Small Acquisition Activity												0.999		0.987**
												(0.008)		(0.005)
Controls														
Proportion of Cross-Border Acquisition.		1.165						1.371		0.859		2.074*		1.018
		(0.266)						(0.347)		(0.227)		(0.775)		(0.232)
Proportion of Cross-Industry Acquisition.		1.724**		1.900**		1.552*						1.181		1.689**
		(0.422)		(0.563)		(0.400)						(0.463)		(0.408)
Proportion of Large Acquisitions		0.432**		0.961		0.406**		0.665		0.515				
		(0.161)		(0.435)		(0.159)		(0.273)		(0.216)				
CEO Overconfidence	1.042	1.051	1.160	1.193	1.137	1.131	1.165	1.171	0.950	0.941	0.899	0.893	1.077	1.082
	(0.119)	(0.121)	(0.152)	(0.158)	(0.136)	(0.137)	(0.144)	(0.145)	(0.123)	(0.122)	(0.163)	(0.167)	(0.125)	(0.126)
CEO Succession	1.092	1.082	1.293*	1.332*	1.082	1.065	1.113	1.109	1.115	1.102	0.944	0.932	1.125	1.137
	(0.152)	(0.152)	(0.199)	(0.206)	(0.157)	(0.155)	(0.166)	(0.166)	(0.172)	(0.171)	(0.209)	(0.209)	(0.158)	(0.160)
Degree of Decentralization	0.717	0.780	0.769	0.767	0.627	0.687	0.694	0.705	1.063	1.126	1.736	1.678	0.688	0.708
	(0.275)	(303)	(0.338)	(0.337)	(0.252)	(0.280)	(0.288)	(0.292)	(0.459)	(0.491)	(1.043)	(1.061)	(0.267)	(0.277)
Performance Above Aspirations	15.186	21.776	0.542	0.508	47.251*	77.910**	205.963**	242.875***	73.127*	47.506*	94.949	274.854*	18.094	19.237
	(30.730)	(44.525)	(1.305)	(1.237)	(98.737)	(164.228)	(437.159)	(517.693)	(169.045)	(111.153)	(299.211)	(888.588)	(36.755)	(39.404)
Performance Below Aspirations	450.238**	323.504**	81.899	52.238	142.939*	97.699*	11.529	12.199	0.251	0.135	130.485	34.713	21.415	16.089
	(1,151.197)	(832.282)	(246.806)	(156.569)	(381.292)	(261.396)	(30.014)	(31.824)	(0.684)	(0.369)	(548.508)	(145.470)	(52.861)	(39.877)
Growth Above Aspirations	0.041***	0.048***	0.116***	0.123***	0.078***	0.090***	0.211***	0.221***	0.107***	0.130***	0.072***	0.098**	0.147***	0.147***
	(0.025)	(0.030)	(0.077)	(0.083)	(0.047)	(0.055)	(0.121)	(0.128)	(0.069)	(0.084)	(0.068)	(0.093)	(0.084)	(0.084)
Growth Below Aspirations	2.366*	2.322*	6.464***	6.873***	1.935	1.873	1.351	1.347	4.537***	4.650***	0.724	0.799	2.922**	3.028**
	(1.147)	(1.145)	(4.048)	(4.368)	(0.966)	(0.959)	(0.675)	(0.684)	(2.637)	(2.715)	(0.492)	(0.550)	(1.467)	(1.528)
Slack Resources	0.206***	0.213***	0.413	0.382	0.154***	0.172***	0.370*	0.394	0.182***	0.162***	0.045***	0.037***	0.401*	0.419
	(0.110)	(0.114)	(0.265)	(0.244)	(0.088)	(0.099)	(0.219)	(0.233)	(0.113)	(0.102)	(0.040)	(0.034)	(0.215)	(0.226)
Acquirer Size	0.646***	0.649***	1.018	1.019	0.679***	0.668***	0.847	0.852	0.674***	0.687***	0.670**	0.593***	0348**	0.771**
	(0.082)	(0.084)	(0.147)	(0.147)	(0.089)	(0.089)	(0.111)	(0.112)	(0.095)	(0.097)	(0.125)	(0.113)	(0.095)	(0.099)
Diversification	1.012	1.050	0.727*	0.683**	0.928	0.975	0.960	0.976	1.111	1.146	1.272	1.486	0.947	0.954
	(0.163)	(0.172)	(0.134)	(0.128)	(0.156)	(0.167)	(0.166)	(0.171)	(0.202)	(0.212)	(0.322)	(0.379)	(0.154)	(0.157)
Industry Acquisition Activity	0.998	1.000	1.003	1.003	0.998	0.999	0.998	0.998	1.004	1.006*	0.998	1.001	1.000	1.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.005)	(0.003)	(0.003)
Constant	0.087***	0.140***	0.006***	0.018***	0.059***	0.106***	0.031***	0.073***	0.032***	0.029***	0.015***	0.043***	0.055***	0.067***
	(0.046)	(0.082)	(0.004)	(0.013)	(0.032)	(0.065)	(0.019)	(0.045)	(0.023)	(0.021)	(0.014)	(0.042)	(0.030)	(0.039)
Observations	2,267	2,267	2,267	2,267	2,267	2,267	2,267	2,267	2,267	2,267	2,267	2,267	2,267	2,267
Number of Firms	187	187	187	187	187	187	187	187	187	187	187	187	187	187
Mundlak Instruments	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Regression coefficients represent odds ratios. Standard errors in their exponentiated form are presented in parentheses. Odd-numbered (even-numbered) model specifications include control variables only (all variables of interest). Only within-firm effects are reported. Between-firm effects (i.e., Mundlak instruments) and year dummies are included in all models but not reported. Light green cells indicate relationships between momentum and acquisition activity variables within the same target firm category (i.e., hypotheses 1 and 2). Conversely, dark green cells indicate relationships between momentum in a lower-complexity target firm category and acquisition activity in a higher-complexity target firm category (i.e., second part of hypothesis 2a).
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

odds ratio in Model 8 is statistically insignificant (p>.10) (coefficients highlighted in light green in Table 4). In other words, whereas increases in cross-border (large) acquisitions seem to reduce cross-border (large) momentum[41], increases in cross-industry acquisitions appear to not affect their corresponding momentum. Moreover, I cannot corroborate my predicted switching behavior. That is, albeit the odds ratios for higher-complexity acquisitions in Models 6, 10, and 14 exceed a value of one, they remain insignificant (p>.10), indicating the absence of my predicted relationship (coefficients colored in dark green in Table 4). The respective AMEs in Table 5 and graphs in Figure 8 support these findings.

Next, I test hypothesis 2b, which predicts that an increase in a firm's acquisition activity in a lower-complexity target firm category vis-à-vis that firm's past acquisition activity level in that category leads to a discontinuation in acquisition momentum in that category. That is, if firms have established an acquisition stream in a lower-complexity category (e.g., small acquisitions), they have no choice to reduce their activity load other than by, *ceteris paribus*, decreasing the volume of acquisitions in that category as there is no other target firm category to which they can switch to reduce their activity load. In other words, acquirers with such acquisition patterns can only alleviate the strains of information overload by discontinuing their momentum. This prediction is supported by statistically significant and smaller-than-one odds ratios in Model 6 (p<.05), Model 10 (p<.01), and Model 14 (p<.05) (coefficients colored in light green in Table 4). Based on these results, a one-unit increase in the number of

[41] The substantially smaller odds ratio for *large acquisition activity* in Model 12 (odds ratio = 0.747) in Table 4 and the 2.2 pp decrease in the predicted probability of *large acquisition momentum* for every additional large acquisition indicate the relative rarity of this type of acquisition.

Table 5: Tests of AMEs for a One-Unit Increase in Acquisition Activity - Baseline Sample

	Total Acquisition Momentum	Higher-Complexity Acquisition Momentum	Lower-Complexity Acquisition Momentum
Total Acquisition Activity	-0.003**		
	(0.001)		
Cross-Border Acquisition Activity		-0.006***	0.003
		(0.002)	(0.002)
Domestic Acquisition Activity		0.004**	-0.004**
		(0.002)	(0.002)
Cross-Industry Acquisition Activity		-0.001	0.001
		(0.001)	(0.001)
Within-Own-Industry Acquisition Activity		0.003	-0.012***
		(0.003)	(0.003)
Large Acquisition Activity		-0.022***	0.013
		(0.006)	(0.014)
Small Acquisition Activity		-0.000	-0.003**
		(0.001)	(0.001)

Notes: The baseline sample comprises N=2,267 firm years, covering 187 unique firms. Standard errors are in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$, two-tailed tests.

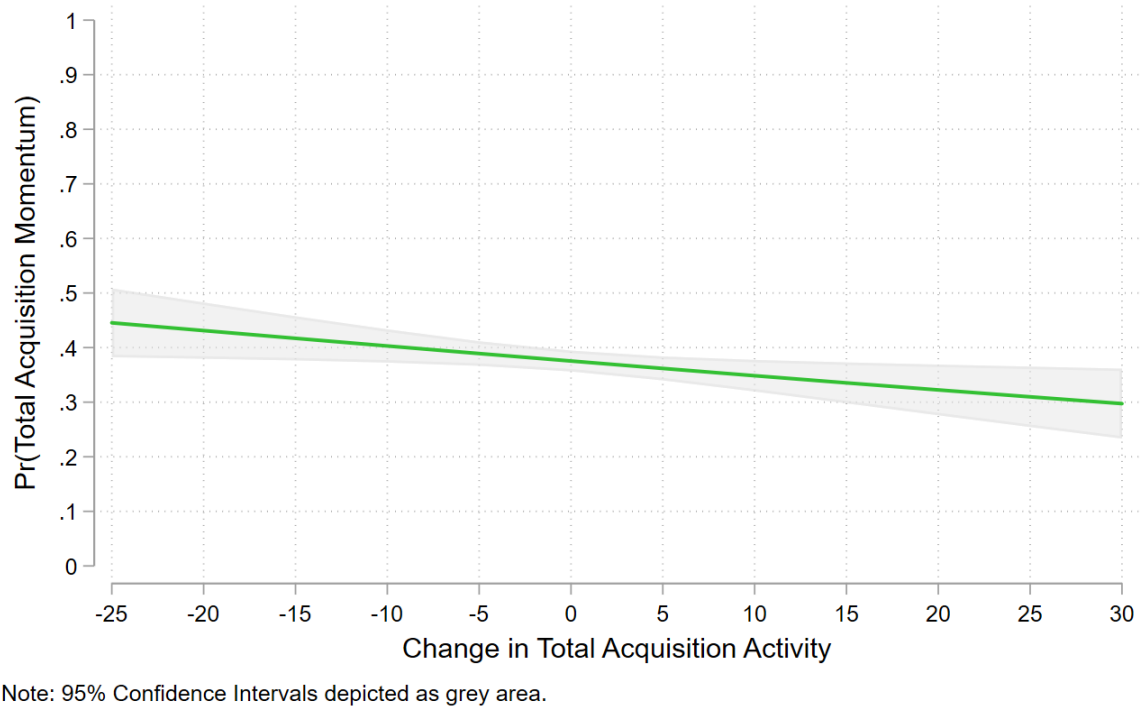


Figure 7: Predicted Probability of Total Acquisition Momentum by Change in Total Acquisition Activity: Main Effect of Total Acquisition Activity

domestic (within-own-industry; small) acquisitions relative to a firm's past acquisition level, on average, corresponds to a 0.4 (1.2; 0.3) pp decrease in the predicted probability of a firm's *domestic (within-own-industry; small) acquisition momentum*. This can also be seen graphically in Figure 8, *which* supports the conclusion that increases in activity load reduce a firm's ability to maintain its momentum.

Furthermore, Table 4 reveals a seemingly surprising relationship between *domestic acquisition activity* and *cross-border acquisition momentum*. More specifically, Model 4

in Table 4 shows that the odds ratio of *domestic acquisition activity* is significant ($p<.05$) and *larger* than one, indicating a 'reverse switch' (i.e., a switch from a lower-complexity category to a higher-complexity category). While this result seems unintuitive from an activity load viewpoint (e.g., Castellaneta & Zollo, 2015), a plausible explanation for it can be found in the international management literature (e.g., Lasserre, 2003; Vermeulen & Barkema, 2001, 2002). That is, as a firm matures, it seeks to expand internationally to benefit from demand in new markets and supply-side cost advantages, among other reasons (e.g., Lasserre, 2003; Vermeulen & Barkema, 2002). One way to achieve this is by acquiring foreign targets, especially if no suitable target can be found in a firm's domestic market (Lasserre, 2003; Vermeulen & Barkema, 2002). Therefore, the positive relationship between *domestic acquisition momentum* and *cross-border acquisition momentum* likely represents the internationalization behavior of the firms in my sample – a behavior which should not surprise as my sample consists of Fortune Global 500 firms, which operate globally.

To explore heterogeneity in acquirer responses due to differences in the absorptive capacity of firms (i.e., hypotheses 3 and 4), I modify my estimation procedure in two ways. First, to assess hypothesis 3 (4), I split my baseline sample into the MECE subgroups *Large Firms* (*Decentralized Firms*) and *Small Firms* (*Centralized Firms*). Doing so, I follow a call of Hoetker (2007), who has argued that fitting separate models for each group should be preferred over interacting my variable of interest with an indicator variable for a firm's group membership when examining cross-group differences in non-linear models. Two facts support his claim: (a) The odds ratio coefficients and significance levels of interaction terms in binary choice models are *observation specific*, rendering them uninformative in their raw form[42] (Ai & Norton, 2003; Hoetker, 2007); (b) using an interaction term and estimating the subsequent binary choice model for *all observations* assumes the unobserved variation for all subgroups to be identical – a strong assumption that produces incorrect estimates if violated (Hoetker, 2007). I thus construct two subsamples using the median of the *Acquirer Size* (*Degree of Decentralization*) distribution in a given year as the cut-off value, creating subgroups that are of roughly equal size. Second, using Mize et al.'s (2019) general framework for comparing effects across non-linear models, I estimate a Generalized Structural Equation Model (GSEM) with the logistic distribution as the link function.[43] Unlike a standard logit model, the GSEM allows me to correctly estimate cross-group differences across separately fitted models, whose calculation requires an estimate of the covariance between the activity load estimates of both subsamples (Mize et al., 2019). In fact, estimating this covariance through a GSEM is crucial as observations from non-overlapping samples do *not* always have a cross-model covariance of zero (Mize et al., 2019). Together, these modifications allow me to test my hypothesized cross-group differences and express these as predicted probabilities (Mize, 2019; Mize et al., 2019).

Recalling hypothesis 3, I expect larger acquirers to respond less strongly to changes in activity load than smaller acquirers. This is because larger acquirers can access a larger pool of managerial resources and more specialized internal acquisition processes (Laamanen & Keil, 2008) that expand their absorptive capacity. A higher absorptive capacity, in turn, alleviates the strains imposed by activity load and, thus, helps larger acquirers maintain their momentum for longer. To corroborate my prediction, I estimate the AMEs of each subsample (i.e., test of first differences) and test for their equality across both groups (i.e., test of second differences). Table 6 reports the results of these tests for a one-unit increase in acquisition activity. Unfortunately, albeit most cross-group differences show the signs predicted by hypothesis 3[44], almost all are not significant ($p>.10$). For instance, the tests in the left column of Table 6 show that a one-unit increase in *total acquisition activity*, on average, is associated with a 0.3 pp and 0.9 pp decrease in the predicted probability of an acquirer's *total acquisition momentum* for large firms and small firms, respectively ($p<.05$ for both AME first differences). However, the insignificant second difference (-0.003 – -0.009 = 0.006; $p>.10$) reveals that the effect of a one-unit increase in activity load, on average, does not differ across groups. That is, larger acquirers, on average, do *not* respond differently to increases in activity load than smaller ones, providing no support for my predictions. The same holds for all other acquisition activity and momentum definitions shown in Table 6, with the significant cross-group effect of *domestic acquisition activity* in the *domestic acquisition momentum* model (-0.003 – -0.017 = 0.013; $p<.05$) presenting an anomaly.

However, although no acquirer size differences in responses to activity load, *on average*, exist across the sample, the non-linear nature of my logit model implies that acquirer size differences can exist *at specific values* (or across a range of values) of activity load (Mize, 2019). A plot of the marginal effects of both subgroups is therefore warranted to identify if and where significant firm size differences exist

[42] I do not report odds ratios for these tests in this section as they provide little meaningful information about interaction effects (Ai & Norton, 2003; Mize, 2019; Mize et al., 2019). In fact, comparing odds ratio coefficients between samples with non-overlapping observations – as one would do in OLS regression – is inappropriate for logit models "because a change in the size of the coefficient across models can reflect both confounding and rescaling of the model [(Karlson et al., 2012)]" (Mize et al., 2019, p. 162). I thus compare my subsamples in the natural metric of my dependent variable: Predicted probabilities, which can be compared across subsamples (Breen et al., 2018; Mize, 2019; Mize et al., 2019). However, for completeness, odds ratio results are presented in Table A.5 and Table A.6 in the appendix.

[43] Note that my overall model specification remains unchanged. That is, I still use a hybrid logit model, and my theoretical interest lies in the within-firm effect of activity load. The only difference is the simultaneous estimation of models that are separately fitted to each subsample.

[44] That is, I expect a positive (negative) cross-group difference in models that regress acquisition momentum on acquisition activity of the same (of a different) complexity level. *Cross-industry acquisition activity* is the only activity load definition which does not show this behavior.

Table 6: Tests of AMEs and Second Differences for a One-Unit Increase in Acquisition Activity - Small Firm and Large Firm Subsamples

	Total Acquisition Momentum			**Higher-Complexity Acquisition Momentum**			**Lower-Complexity Acquisition Momentum**		
	AME_{Small}	AME_{Large}	Second Difference	AME_{Small}	AME_{Large}	Second Difference	AME_{Small}	AME_{Large}	Second Difference
Total Acquisition Activity	-0.009**	-0.003**	0.006						
	(0.004)	(0.001)	(0.004)						
Cross-Border Acquisition Activity				-0.009	-0.006***	0.003	0.003	0.002	-0.001
				(0.006)	(0.002)	(0.006)	(0.007)	(0.002)	(0.008)
Domestic Acquisition Activity				0.008	0.004*	-0.004	-0.017***	-0.003*	0.013**
				(0.006)	(0.002)	(0.006)	(0.005)	(0.002)	(0.006)
Cross-Industry, Acquisition Activity				0.000	-0.001	-0.002	-0.002	0.001	0.003
				(0.005)	(0.001)	(0.005)	(0.005)	(0.002)	(0.005)
Within-Own-Industry Acquisition Activity				0.006	0.001	-0.005	-0.021***	-0.011***	0.009
				(0.008)	(0.003)	(0.009)	(0.006)	(0.003)	(0.007)
Large Acquisition Activity				-0.037***	-0.016*	0.020	0.032	-0.000	-0.032
				(0.011)	(0.008)	(0.014)	(0.025)	(0.016)	(0.030)
Small Acquisition Activity				*0.001*	*-0.000*	*-0.001*	-0.004	-0.003**	0.001
				(0.003)	(0.001)	(0.003)	(0.004)	(0.001)	(0.004)

Notes: The small firm (large firm) subsample comprises N=1,138 (N=1,129) firm years, covering 112 (131) unique firms. Standard errors are in parentheses.
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$, two-tailed tests.

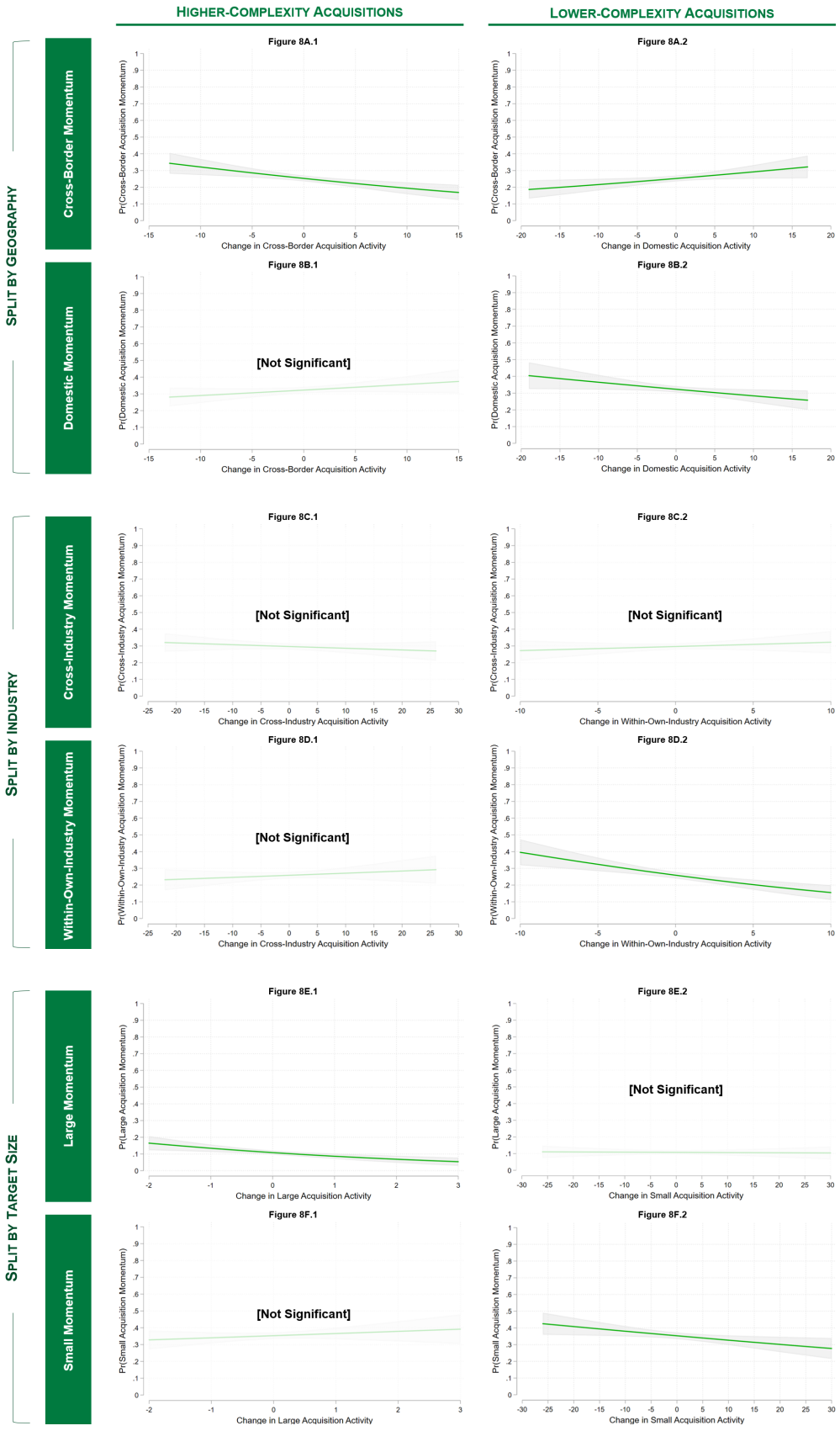


Figure 8: Predicted Probability of Dimension-Specific Acquisition Momentum by Change in Dimension-Specific Acquisition Activity: Heterogeneity in Acquirer Responses due to Differences in Acquisition Complexity

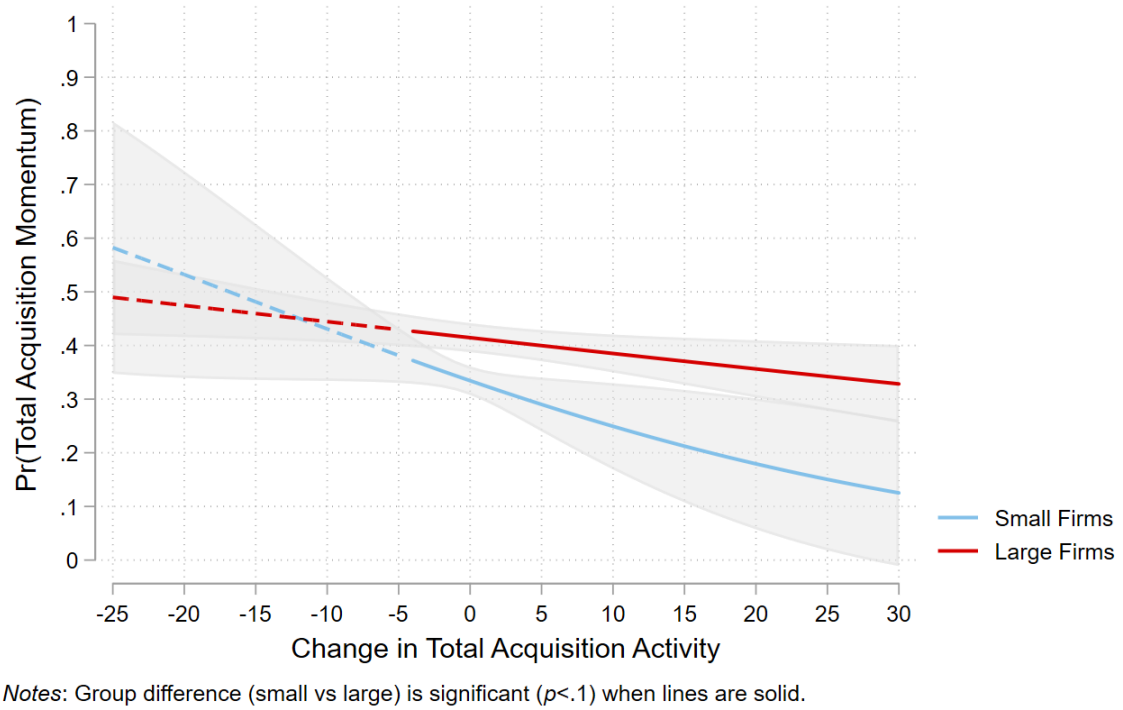


Figure 9: Predicted Probability of Total Acquisition Momentum by Acquirer Size and Change in Total Acquisition Activity: Interaction Effect Between Acquirer Size and Total Acquisition Activity

across the range of acquisition activity values in my sample (Hoetker, 2007; Mize, 2019). Figure 9 presents the predicted probability of large and small acquirers to maintain their *total acquisition momentum* across the range of *total acquisition activity*. A solid line indicates intervals with significant ($p<.10$) acquirer size differences. Specifically, Figure 9 shows that acquirer size differences are significant — with large acquirers having a higher likelihood of maintaining their momentum — when comparing firms that reduce the number of acquisitions by less than five acquisitions or even increase it relative to their past acquisition activity level (all contrasts $p<.10$). In other words, larger acquirers are more likely to maintain their acquisition momentum in the face of increases in activity load due to their relative resource abundance. Intuitively, this *level difference* in the probability of maintaining momentum seems to grow between both subgroups as the activity load from acquisitions increases.[45] Conversely, no cross-group differences in the probability of maintaining momentum exist when acquirers reduce the number of acquisitions by five or more relative to their past acquisition activity level (all contrasts $p>.10$). A detailed visual inspection of the *slopes* of both groups further indicates that cross-group differences in responses to a one-unit increase in *total acquisition activity* may exist *at specific values* of my independent variable.[46] This pattern is mostly robust to changes in acquisition activity and momentum definitions as shown in Figure 10. Overall, these results reveal that larger acquirers can bear a higher activity load burden – and possibly respond less strongly to changes in activity load – for certain levels of acquisition activity due to their relative resource abundance, partially confirming hypothesis 3 and its underlying Penrosian logic.

Finally, in hypothesis 4, I expect acquirers with a more decentralized organizational structure to respond less strongly to changes in activity load relative to acquirers with a more centralized organizational structure. That is, a higher degree of decentralization allows firms to distribute the activity load from acquisitions more evenly across their resources, avoiding a too strong concentration of activity load within one single part of the organization. Like an increase in firm size, a more decentralized organizational structure thus expands the absorptive capacity of acquirers, making acquirers with such a structure more likely to maintain their momentum. Table 7 presents tests of subsample AMEs and second differences for a one-unit increase in acquisition activity, revealing that almost all cross-group differences are insignificant ($p>.10$) and rarely show the effect directions predicted by hypothesis 4.[47] For instance, the tests in the left column of Table 7 indicate that the AME of a one-unit increase in *total acquisition activity*, on average, corresponds to a 0.3 pp decrease in the probability of an acquirer's *total acqui-*

[45] That is, small acquirers are disproportionately strongly affected by large increases in acquisition activity due to their relative resource scarcity.

[46] I refrained from testing second differences at specific values of acquisition activity for their significance due to time and space constraints.

[47] That is, as I did for acquirer size differences, I expect a positive (negative) acquirer structure difference in models that regress acquisition momentum on acquisition activity of the same (of a different) complexity level.

Table 7: Tests of AMEs and Second Differences for a One-Unit Increase in Acquisition Activity - Centralized Firm and Decentralized Firm Subsamples

	Total Acquisition Momentum			**Higher-Complexity Acquisition Momentum**			**Lower-Complexity Acquisition Momentum**		
	$\text{AME}_{Central}$	$AME_{Decentral}$	Second Difference	$\text{AME}_{Central}$	$AME_{Decentral}$	Second Difference	$\text{AME}_{Central}$	$AME_{Decentral}$	Second Difference
Total Acquisition Activity	-0.003*	-0.003	0.000						
	(0.002)	(0.002)	(0.002)						
Cross-Border Acquisition Activity				-0.005**	-0.008***	-0.003	0.005	0.000	-0.004
				(0.002)	(0.003)	(0.004)	(0.003)	(0.004)	(0.005)
Domestic Acquisition Activity				0.001	0.006**	0.005	-0.005*	-0.003	0.002
				(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.004)
Cross-Industry Acquisition Activity				-0.002	-0.001	0.001	0.001	0.002	0.001
				(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
Within-Own-Industry Acquisition Activity				0.000	0.004	0.004	-0.011***	-0.012***	-0.001
				(0.004)	(0.005)	(0.006)	(0.004)	(0.004)	(0.006)
Large Acquisition Activity				-0.025**	-0.026***	-0.001	0.027	-0.002	-0.029
				(0.012)	(0.009)	(0.015)	(0.021)	(0.018)	(0.028)
Small Acquisition Activity				-0.003*	0.003***	0.006***	-0.003	-0.004*	-0.001
				(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)

Notes : The centralized firm (decentralized firm) subsample comprises N=991 (N=1,276) firm years, covering 144 (168) unique funs. Standard errors are in parentheses.
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$, two-tailed tests.

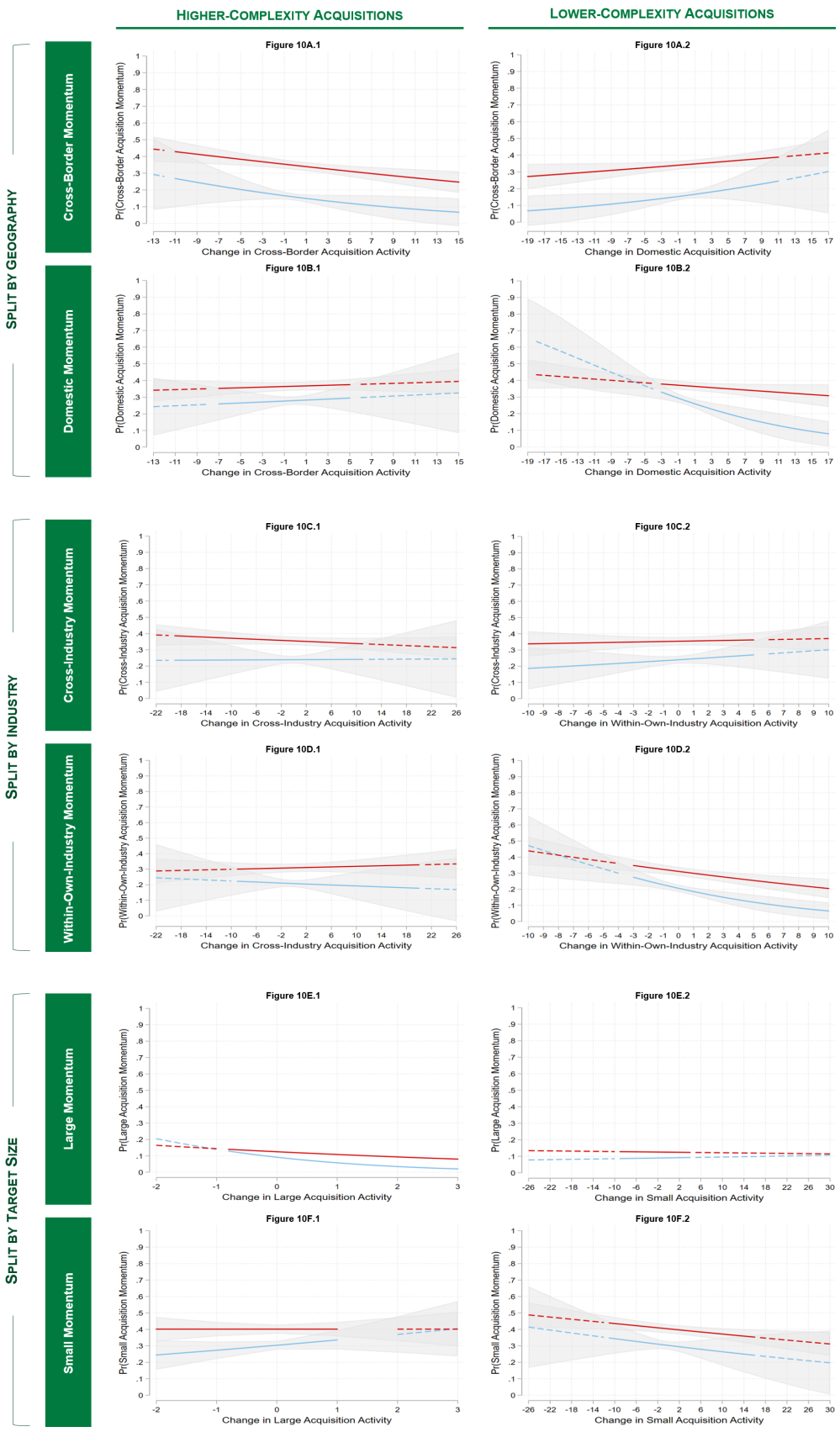
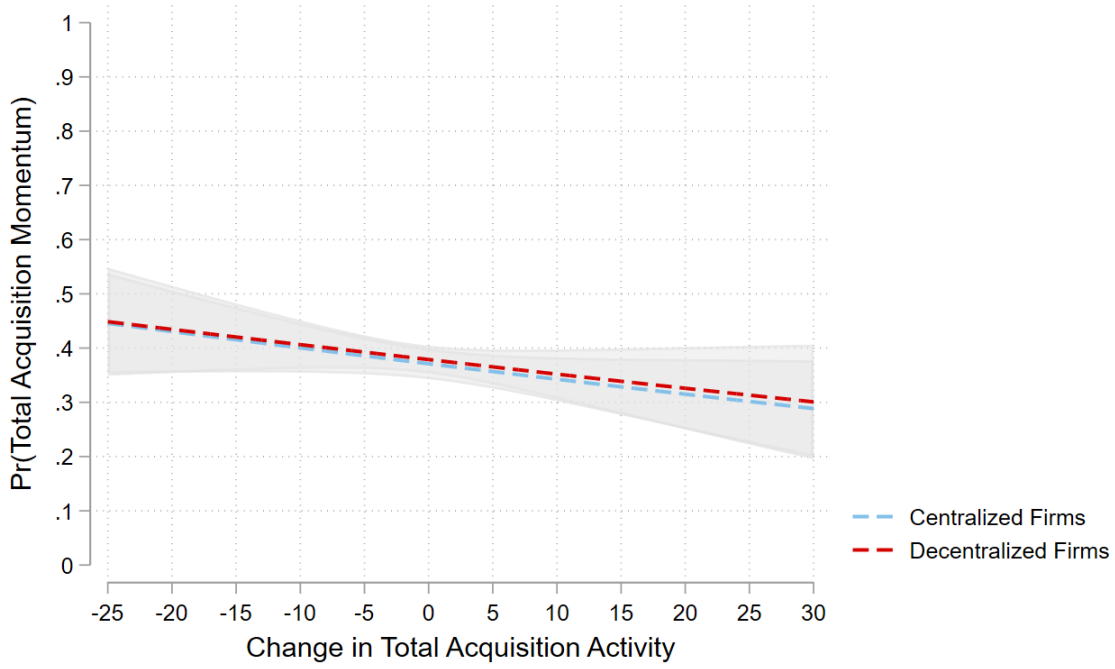


Figure 10: Predicted Probability of Dimension-Specific Acquisition Momentum by Acquirer Size and Change in Dimension-Specific Acquisition Activity: Heterogeneity in Acquirer Responses due to Differences in Acquirer Size and Acquisition Complexity



Notes: Group difference (central vs decentral) is significant (p<.1) when lines are solid.

Figure 11: Predicted Probability of Total Acquisition Momentum by Acquirer Structure and Change in Total Acquisition Activity: Interaction Effect Between Acquirer Structure and Total Acquisition Activity

sition momentum for centralized firms (p<.10), whereas the AME for decentralized firms is insignificant (p>.10). Similarly, the insignificant second difference (-0.003 – -0.003 = 0.000; p>.10) shows that the effect of a one-unit increase in activity load, on average, does not differ across groups. In other words, firms with a decentralized structure, on average, do *not* respond differently to increases in activity load than their centralized counterparts, providing no support for my predictions. This result holds consistently across the acquisition activity and momentum definitions shown in Table 7, with the significant cross-group effect of *small acquisition activity* in the *large acquisition momentum* model (0.003 – -0.003 = 0.006; p<.01) being an outlier.

As I did before, I plot the marginal effects of both groups to test whether cross-group differences exist at specific values of my independent variable (Mize, 2019). Figure 11 shows the predicted probability of decentralized and centralized firms to maintain their *total acquisition momentum* across the range of *total acquisition activity*. Interestingly, Figure 11 reveals that no significant *level differences* in the probability of maintaining momentum exist between decentralized and centralized acquirers (all contrasts p>.10). That is, decentralized acquirers are *not* more likely to maintain their acquisition momentum than centralized acquirers. Furthermore, given that both curves look almost identical *slope-wise*, cross-group differences in responses to a one-unit increase in *total acquisition activity* appear to not exist across all values of my independent variable. Substantively, this indicates that decentralized acquirers do not respond less strongly to changes in acquisition activity, rejecting hypothesis 4. This pattern is mostly robust to changes in acquisition activity and momentum definitions as shown in Figure 12. In sum, these findings indicate that an acquirer's degree of decentralization barely affects that firm's absorptive capacity and, thus, does not explain heterogeneity in firm responses to activity load.

4.3. Robustness Tests

There are four major concerns about the results from my main analysis. First, one could doubt whether the selected acquisitions in my baseline sample accurately reflect the intuition behind my stipulated activity load mechanism. That is, while all acquisitions are resource-consuming endeavors, minority-stake acquisitions appear to be less likely to cause an information overload due to the absence of post-merger integration activities. In fact, such acquisitions are frequently treated as financial investments and, thus, likely consume less resources than acquisitions that require the integration of the target organization. Second, it could be argued that my baseline operationalization of momentum does not account for the extended multi-year time horizon that is normally associated with acquisition streams (e.g., Laamanen & Keil, 2008). In other words, my current operationalization, which defines momentum as the change in known deal values from period *t-1* to *t*, appears to examine a time horizon that is too short to make inferences about the long-term acquisition behavior of firms. Third, one could question whether suitable proxies are used for the acquisition complexity dimensions in my main analysis – a concern that applies to all

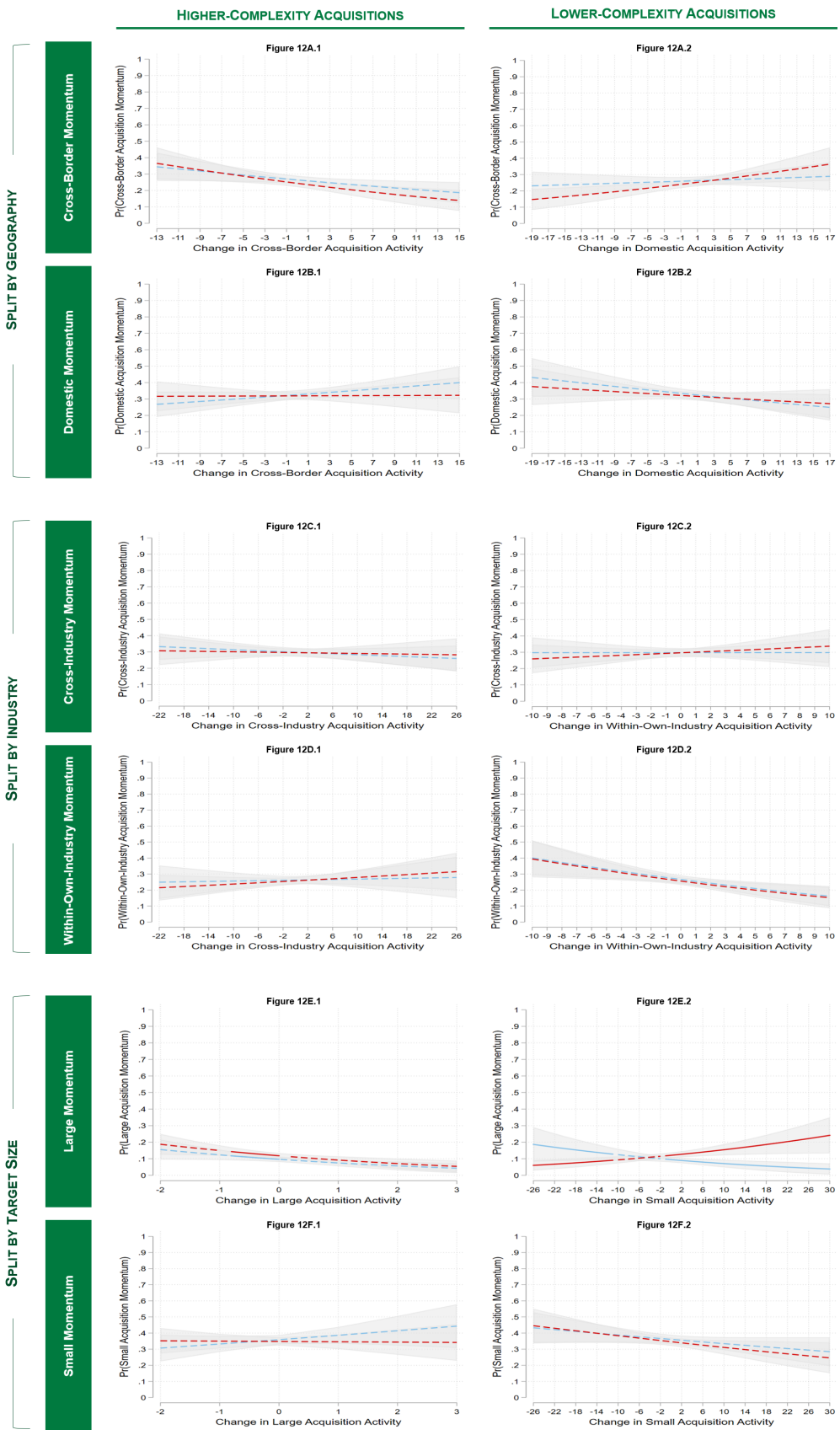


Figure 12: Predicted Probability of Dimension-Specific Acquisition Momentum by Acquirer Structure and Change in Dimension-Specific Acquisition Activity: Heterogeneity in Acquirer Responses due to Differences in Acquirer Structure and Acquisition Complexity

aggregation levels shown in Figure 6.[48] Fourth, one could have reservations about the construct validity of my variable *degree of decentralization*. Specifically, it could be criticized that executive titles may not accurately proxy structural features of acquirers and, thus, produce incorrect results in my analysis of heterogeneity in acquirer responses.

To eliminate these concerns, I adapt the models in my main analysis in four ways. First, I re-run my main analysis with two modified samples, of which the first only covers majority-stake acquisitions and the second comprises majority-stake acquisitions and acquisitions for which no acquired stake was reported in the Refinitiv database. Second, to address concerns about my baseline operationalization of momentum, I adapt the definition of momentum in two ways. On the one hand, I broaden the time window of my variable from two consecutive one-year periods (i.e., *t-1* to *t*) to two consecutive two-year and three-year periods. That is, I define momentum as the change in known deal values from periods *t-2* and *t-1* (*t-3*, *t-2*, and *t-1*) to *t* and *t+1* (*t*, *t+1*, and *t+2*). On the other hand, I change the operationalization of momentum by replacing monetary deal values with an acquirer's number of acquisitions, using the three different time windows from before to discern effects of alternative time structures. Third, I operationalize my acquisition complexity dimensions differently. For instance, I account for the particularly strong homogeneity between firms from Anglo-Saxon countries (Ronen & Shenkar, 2013) by distinguishing between cross-region and intra-region acquisitions, with 'intra-region' denoting acquisitions of targets which are headquartered in Australia, Canada, Ireland, New Zealand, the United Kingdom, or the United States. Similarly, I define a firm's industry based on the first three digits of that firm's SIC code and distinguish between large and small acquisitions through an acquirer-specific mean of the target-to-acquirer-total-assets ratio, which I compute based on my baseline sample. Fourth, to ensure the validity of my variable *Degree of Decentralization*, I re-run my analysis of heterogeneity in acquirer responses with an operationalization that uses an acquirer's number of business units as a proxy for that firm's organizational structure.[49]

Table A.7 in the appendix presents results of my first set of robustness tests, which use samples that exclude minority-stake acquisitions from the analysis.[50] Overall, it seems that my results are very robust to the exclusion of minority-stake acquisitions – both in terms of significance levels and effect sizes. Yet, two findings should be noted. First, in Model 5 and Model 6, the effect of *domestic acquisition activity* on *domestic acquisition momentum* in my main analysis appears to be driven by minority-stake acquisitions. This is indicated by odds ratio coefficients of *domestic acquisition activity* that are less significant ($p<.10$) or insignificant ($p>.10$) and have values closer to one relative to those in my main analysis. Second, in Models 11 and 12, it seems that acquisitions for which no stake was disclosed drive the effect of *large acquisition activity* on *large acquisition momentum* in my main analysis. Like before, this is shown by odds ratio coefficients of *large acquisition activity* that have values closer to zero (one) and are more significant ($p<.01$) (insignificant ($p>.10$)) in Model 11 (12) than those in my main analysis.

The results of my second set of robustness tests, which assess different operationalizations of momentum, are shown in Table A.8 and Table A.9 in the appendix. In sum, it appears that my results are very robust to changes in the time structure of momentum as shown in Table A.8. A closer inspection of Table A.8, however, reveals two interesting findings. First, almost all model specifications have odds ratio coefficients that are more extreme in magnitude[51] compared to those in my main analysis, and the extremity of this magnitude increases with the length of the time window used. In other words, the longer the time window that is used to define momentum, the more pronounced the effect of activity load. Second, unlike in my main analysis, increases in *cross-industry acquisition activity* reduce an acquirer's likelihood of pursuing further cross-industry acquisitions, providing further support for hypothesis 2a. This is indicated in Model 8 by a significant ($p<.05$) and lower-than-one odds ratio coefficient of *cross-industry acquisition activity*. The results presented in Table A.9, in which monetary deal values are replaced by an acquirer's number of acquisitions, further solidify these conclusions. That is, my hypothesized effects are stronger and often more significant in Table A.9 than in my main analysis, and these effect sizes increase with the length of the time window used for defining momentum. In addition, almost all relationships predicted in hypotheses 1, 2a, and 2b are highly significant ($p<.01$) in model specifications that use a three-year-on-three-year time structure in their definition of momentum.[52] The consistency of this finding across different proxy variables in Table A.8 and Table A.9

[48] For instance, my distinction between cross-border and domestic acquisitions abstracts from structural similarities between countries in the same regional cluster (e.g., Ronen & Shenkar, 2013) and, thus, might measure complexity differences between acquisitions in different geographies in an overly simplistic way. Likewise, the first two digits of a target's SIC code might not sufficiently reflect homogeneity of acquisitions on the industry level as targets in the same major group (i.e., with identical first two digits in their SIC code) exhibit lower levels of similarity than targets in the same industry group (i.e., with identical first three digits in their SIC code), potentially pooling non-homogenous acquisitions within the same complexity cluster. Also, my distinction between large and small acquisitions, which is based on the sample mean of the target-to-acquirer-total-assets ratio, may be too one dimensional as the relative size of an acquisition likely is acquirer specific.

[49] Based on Compustat business segment data.

[50] Due to space constraints, the results of my robustness tests are presented in multiple tables in the appendix.

[51] That is, odds ratio coefficients have values that are closer to zero or infinity, depending on the hypothesized effect direction.

[52] More specifically, all predicted relationships that were supported in my main analysis also hold in Models 3, 6, 9, 15, 18 and 21 in Table A.9. However, unlike in my main analysis, the effect of *cross-industry acquisition activity* is significant ($p<.01$) in Model 12. Combining this observation with the positive and significant ($p<.01$) odds ratio coefficient of *cross-border acquisition activity* (*cross-industry acquisition activity*) in Model 9 (15), I find support for the 'switching' behavior from higher-complexity targets to lower-complexity targets as predicted in hypothesis 2a. In fact, only the odds ratio coefficient of *large acquisition activity* in Model 21 remains insignificant ($p>.10$), implying that no switch from

implies that momentum definitions that use longer time windows likely are superior operationalizations in the context of research questions that explore phenomena which unfold over extended periods, such as acquisition patterns.

Table A.10 in the appendix illustrates results of my third set of robustness tests, which explore alternative operationalizations of my acquisition complexity dimensions. Overall, it seems that my results are mostly robust to definitory changes in my complexity dimensions – both in terms of effect sizes and significance levels. Two exceptions should be noted, however. First, the effect of *intra-region acquisition activity* in Model 3 is insignificant ($p>.10$). This indicates that my results are not robust to changes in the definition of country-level complexity, possibly because the Anglo-Saxon region is a region with high acquisition activity[53] that allows firms to maintain their momentum by acquiring targets in other countries within that region if targets in one specific country become exhausted. Second, unlike in my main analysis, the effect of *small acquisition activity* is significant ($p<.10$) in Model 6. Although this finding is counterintuitive, a closer look at the regression output reveals that the odds ratio coefficient of *small acquisition momentum* is marginally significant[54], indicating that the substantive importance of this finding is limited. Thus, my results are rather robust to definitory changes in all complexity dimensions.

Finally, Table A.11 in the appendix presents results of my last set of robustness tests, which use an acquirer's number of business units as a proxy for that firm's *Degree of Decentralization*. In sum, it seems that my results are *not* robust to alternative operationalizations of a firm's *Degree of Decentralization*. Specifically, I make two observations in Table A.11. First, the robustness of my results varies with the dependent variable. For instance, in most model specifications, the main effects in my subsample of centralized (decentralized) firms are slightly stronger (weaker) than those in main analysis, which are illustrated in Table A.6 in the appendix. This is indicated by slightly more (less) extreme odds ratio coefficients for centralized (decentralized) firms in Models 1 to 4 and Models 9 to 12 in Table A.11. Conversely, the odds ratio coefficients in Models 5, 6, 13, and 14 show the opposite behavior with varying magnitudes and unsystematic changes in their significance levels. This implies that model specifications in which *domestic acquisition momentum* or *small acquisition momentum* are the dependent variable show lower levels of robustness than models with other dependent variables. Second, across almost all model specifications, it seems that decentralized firms, on average, *might* respond less strongly than centralized ones. This is indicated by odds ratio coefficients that have values closer to one for decentralized firms in Models 2, 4, 10, 12, and 14. This pattern, if corroborated by tests of second differences, would support my predictions in hypothesis 4 and, thus, indicate that a firm's number of business units likely proxies structural features of firms more accurately.

large to small acquisitions occurs – a finding that could indicate that target size may not be a relevant complexity dimension.

[53] Almost 70 percent (= 8,311 / 11,951) of acquisitions in my baseline sample occur in the Anglo-Saxon regional cluster as defined by Ronen and Shenkar (2013). Please see Figure A.3 in the appendix for a visual breakdown of all acquisitions by their country of origin.

[54] More specifically, this odds ratio coefficient has a p-value of $p=.087$.

5. Discussion

In this section, I put my empirical findings into perspective by benchmarking them against evidence in extant literature. For this, I resort to the literature streams introduced in section 2, highlighting the implications of my results for the literature on temporal acquisition patterns, strategic momentum, and activity load. Moreover, acknowledging the high practical relevance of acquisitions, I discuss the practical implications of my findings for managers. Finally, this section presents the limitations of my study as well as potential alleys for future research.

5.1. Review and Implications of Empirical Findings

In this thesis, I aim to contribute to our understanding of the factors that cause acquiring firms to systematically deviate from their established, externally observable acquisition patterns. For this, I explore the role of a factor that has not yet been studied in acquisition pattern research: Activity load. Building on this construct, I find empirical support for my baseline hypothesis, which predicts that increases in a firm's activity load from acquisitions are negatively related to that firm's likelihood of pursuing future acquisitions. In other words, increases in the volume of acquisitions induce acquirers to decelerate their acquisition pace to alleviate the strains on their resources imposed by high levels of activity load. However, I cannot find robust support for my second predicted acquirer response: A switch from acquisitions in a higher-complexity target firm category to acquisitions in a lower-complexity target firm category. That is, firms may or may not switch to target firms that are less complex to acquire – and, thus, change the observable structural properties of their acquisition pattern – to reduce their activity load from acquisitions, depending on the operationalization of momentum. Likewise, I can only partially corroborate that structural features of acquirers, such as their size or organizational structure, explain differences in the acquisition behavior of these firms. Specifically, while an acquirer's size seems to be related to that firm's capacity to absorb acquisitions, I find no robust support for the moderating effect of a firm's organizational structure. Thus, it seems that the availability of resources influences a firm's absorptive capacity and, through this, its acquisition behavior, but the precise nature of this moderating relationship has yet to be fully understood.

These findings have important implications for scholars and practitioners alike:

5.1.1. Contributions to Research on Temporal Acquisition Patterns

This thesis contributes to extant research on temporal acquisition patterns in two ways. First, by exploring whether

a firm's activity load from acquisitions can explain deviations from stable acquisition patterns, this thesis adds to our understanding of the antecedents of these patterns. More specifically, it complements prior studies that have largely focused on the performance implications of acquisition patterns (e.g., Ellis et al., 2011; Hayward, 2002; Laamanen & Keil, 2008; Schipper & Thompson, 1983; Shi & Prescott, 2011; Vermeulen & Barkema, 2002) or explored factors that solely explain differences in acquisitiveness between firms (e.g., Baum et al., 2000; Bertrand et al., 2019; Haleblian et al., 2017; Kim et al., 2011; Malmendier & Tate, 2008; Shi et al., 2017). Closing this research gap was essential given the detrimental firm-level performance effects associated with systematic deviations from previously stable acquisition patterns (Ellis et al., 2011; Laamanen & Keil, 2008; Vermeulen & Barkema, 2002). Second, this thesis offers new insights to scholars who identify acquisition patterns based on externally observable target firm attributes (e.g., Ellis et al., 2011; Hayward, 2002). That is, while prior studies have explored the timing of acquisitions in acquisition patterns (e.g., Hayward, 2002) and the transferability of learnings across targets with different observable attributes (e.g., Ellis et al., 2011), this thesis assesses structural changes within acquisition patterns that unfold over time. Specifically, my robustness tests reveal that such structural changes *potentially* exist and manifest themselves in a switching response – from targets in a higher-complexity category to targets in a lower-complexity category – that allows acquirers to reduce their activity load while maintaining their overall momentum. In fact, it almost seems as if high levels of activity load flip a switch within the organization of acquirers that induces these firms to deviate from their established acquisition patterns. This novel insight, combined with a mechanism that draws on Penrosian resource-based logic and the BTF (e.g., Cyert & March, 1963; March & Simon, 1958; Simon, 1945), adds to prior research, which has largely investigated this phenomenon from an organizational learning perspective, and thus furthers our understanding of acquisition patterns.

5.1.2. Contributions to Research on Strategic Momentum

Besides adding to our understanding of temporal acquisition patterns, my findings make two contributions of theoretical nature to the debate on strategic momentum. On the one hand, my results warrant the separation of within- and between-firm effects in momentum research – both in theory development and empirical measurement. This need arises as my results confirm prior empirical findings of Beck et al. (2008), who have discovered that firms show behavior *opposite* to the repetitive momentum hypothesis once unobserved time-invariant heterogeneity on the firm level is accounted for (i.e., once the within-firm effect is fully isolated). Substantively, this indicates not only that the repetitive momentum hypothesis in its original formulation does not hold but also that past studies of strategic momentum that used random effects models, which do not separate within- and between-firm effects, have likely reported biased results. Separating within- and between-firm effects thus seems warranted to mitigate any biases and allow for a more fine-grained theory development in momentum research. On the other hand, my stipulated causal mechanism expands the theoretical underpinnings of prior momentum research. That is, whereas past studies have drawn on organizational learning theory and theories of managerial cognition to explain their findings, this thesis uses Penrosian resource-based logic as a novel theoretical lens. Doing so, I corroborate that the finite cognitive capacity of managers limits a firm's ability to maintain momentum, going beyond the mechanisms of routine formation and cognitive maps that were heavily emphasized in prior research (e.g., Amburgey & Miner, 1992).

Additionally, this thesis contributes one methodological insight to extant research on strategic momentum. That is, my robustness tests indicate that the length of the time window used in the operationalization of momentum strongly affects the effect size and significance level of the effect of activity load. Although this finding may surprise at first glance, it can be explained by a key characteristic of my phenomenon of interest: Temporal acquisition patterns evolve over time. Thus, meaningful structural changes in these patterns only become visible after a certain period of time has elapsed. Empirically, this implies that longer time intervals of acquisitions need to be contrasted with each other to discern true structural changes. By deriving this insight, this thesis makes an important methodological contribution to extant momentum research, which has predominantly relied on event history analysis (e.g., Amburgey & Miner, 1992).

5.1.3. Contributions to Research on Activity Load and Absorptive Capacity

This thesis also contributes to research on activity load and absorptive capacity in three ways. First, my findings complement prior research on activity load by studying whether activity load can explain the acquisition behavior of firms. That is, while prior research has largely explored the performance implications of (e.g., Kusewitt, 1985; Laamanen & Keil, 2008; Shaver, 2006; Zorn et al., 2019) and structural responses to activity load (e.g., Barkema & Schijven, 2008), this thesis presents evidence on activity load being an antecedent of acquisition behavior, closing a major gap in extant research. Second, this thesis adds to past studies by assessing the joint effect of acquisition volume and complexity. Specifically, my robustness tests reveal that acquirers *potentially* switch from targets in a higher-complexity target firm category to targets in a lower-complexity target firm category to reduce their activity load burden while maintaining their overall momentum. With this insight, my thesis contributes to prior research, which has mostly investigated the effects of acquisition volume and complexity in isolation (e.g., Barkema & Schijven, 2008; Castellaneta & Zollo, 2015; Kusewitt, 1985; Vermeulen & Barkema, 2002; Zorn et al., 2019) or only partially substantiated the existence of a joint effect of acquisition volume and complexity (e.g., Laamanen & Keil, 2008). Third, this study adds to our understanding of the interplay of activity load and absorptive

capacity in two ways. That is, my results confirm results of past studies which have shown that the size of acquirers weakens the negative effect of activity load (Laamanen & Keil, 2008). Thus, it appears that a larger firm size, *ceteris paribus*, helps firms manage the strains of activity load more effectively. In addition, my results provide no robust support for the moderating effect of a firm's organizational structure as predicted by Penrose (1959). This insight is crucial because this thesis is – to the best of my knowledge – the first to empirically test this moderating effect in the context of acquisitions. Yet, this result does not necessarily preclude the existence of the moderating effect of a firm's organizational structure, mainly due to methodological reasons.[55] Taken together, my findings largely corroborate that the concepts of activity load and absorptive capacity represent two sides of the same coin, supporting the rationale behind Penrose's (1959) 'fundamental ratio'.

5.1.4. Managerial Implications

Beyond these contributions to academia, this thesis sensitizes managers to the consequences of high levels of activity load. That is, when making decisions about acquisitions, managers need to be cognizant of the level of activity load already borne by members within the organization. This is because the finite cognitive capacity of these members limits the number of acquisitions a firm can do in a given period (Penrose, 1959). Otherwise, the activity load from acquisitions can lead to information overload on the individual level (Castellaneta & Zollo, 2015) and corporate indigestion on the organizational level (Kusewitt, 1985) – two consequences that are associated with negative performance implications for acquirers (Laamanen & Keil, 2008).

Finally, this thesis provides managers with a set of practical actions. That is, my results suggest that managers *can* actively manage the activity load from acquisitions. Specifically, managers can choose between three options that vary in their ease of implementation. First, managers can consciously regulate the volume of acquisitions to benefit from inorganic growth while avoiding a situation of information overload. While easy to implement, this option may not always be aligned with the strategic goals of the acquirer. Second, if this option is not desired, managers can decide to acquire targets that are less complex to acquire or integrate. However, the feasibility of this option depends on a firm's ability to switch to a lower-complexity target firm category and, thus, may not be available to all acquirers. Third, instead of regulating the activity load of acquisitions, managers can increase the absorptive capacity of their firm. While this can be achieved, for instance, by hiring and training additional people, this option is hard to implement in the short run because the accumulation of knowledge in new hires with respect to M&A processes and specificities of the acquirer organization requires time (Penrose, 1959). Managers therefore need to carefully evaluate these trade-offs in accordance with the strategic goals of and resource base available to their organization before choosing a specific option.

5.2. Limitations and Recommendations for Future Research

My study has two main limitations that may affect the validity of my findings. First, my analysis fully abstracts from a key driver of acquisition behavior: Strategic intent (e.g., Biggadike, 1979; Karim & Mitchell, 2000; Lasserre, 2003; Lee & Lieberman, 2010). In fact, given the multitude of rationales behind acquisitions, the acquisition behavior of firms likely is more opportunistic than assumed in this study. This could raise concerns about unaddressed confounding factors in my analysis and, thus, limit the explanatory power of my results. Second, albeit I extensively draw on prior theory to explain the underlying mechanics of my predicted activity load effect, I did not directly measure my proposed causal mechanism. This implies that my results could reflect a spurious correlation instead of a causal effect. Thus, to prove the validity of my causal mechanism, future research needs to explicitly rule out competing mechanisms, such as changes in a firm's strategic intent (e.g., Biggadike, 1979; Karim & Mitchell, 2000; Lasserre, 2003; Lee & Lieberman, 2010) and a firm's past acquisition performance (e.g., Kusewitt, 1985; Laamanen & Keil, 2008; Zorn et al., 2019). This can, for instance, be achieved through qualitative research designs, which are particularly well suited to examine process-based mechanisms like mine due to "their capacity to capture temporally evolving phenomena in rich detail, something that is hard to do with methodologies . . . [that use archival data]" (Langley & Abdallah, 2011, p. 202).

Furthermore, two methodological choices in my study could limit the explanatory power of my results. First, my chosen sample of acquirers may compromise the generalizability of my results to the full population of acquirers. This is because my sample only includes firms that are large in size and mainly located in the United States. However, while this choice could induce a firm-size and/or geographic bias, it ensures that my study only analyzes acquirers that are sufficiently acquisitive – a pre-condition for examining acquisition patterns that evolve over time. Second, some operationalizations of variables in this study might raise concerns about construct validity. For example, the results of my robustness tests indicate that my baseline operationalization of structural features of firms, which is based on executive titles, could suffer from this problem, potentially causing inconclusive results. Similar concerns could apply to my operationalization of activity load. That is, although my study followed prior research (e.g., Castellaneta & Zollo, 2015) by using externally observable target firm features as proxies for a firm's activity load, my proxies fully abstract from internally observable or unobservable acquisition characteristics that could influence the activity load from acquisitions, such as differences in organizational cultures and the degree of structural integration in the PMI stage (Haspeslagh & Jemison, 1991). Moreover, even if externally observable target

[55] That is, my baseline operationalization of *Degree of Decentralization* might not accurately proxy structural features of acquiring firms.

firm attributes were valid proxies for activity load, the construct validity of my measures can still be questioned. This is because, unlike prior studies that use *relative* proxies, such as target-to-target similarity, to measure heterogeneity in acquisitions (e.g., Hayward, 2002; Laamanen & Keil, 2008), I use *absolute* proxies, such as the number of domestic acquisitions, to account for differences in acquisition complexity. These proxies, however, are rather static and, thus, abstract from the dynamic properties of activity load (Castellaneta & Zollo, 2015), possibly limiting the explanatory power of my findings.

Considering these limitations, future research could add to our understanding of the effect of activity load by exploring four topic areas. First, given the complex nature of the construct of activity load, scholars could examine alternative, more advanced operationalizations of activity load. Future studies could, for instance, re-run my analyses with an acquirer's target-to-target similarity as a measure for heterogeneity in acquisitions. Alternatively, scholars could construct a multi-dimensional index as shown in Figure A.4 in the appendix instead of analyzing possible dimensions of activity load separately, thereby complementing the approach used in this thesis. Second, future research could further explore factors that moderate the effect of activity load. This would be critical since our understanding of variables that moderate the effect of activity load remains limited, both with respect to variables that were explored and not explored in this study. Third, scholars could investigate the relationships between different types of firm-level responses to activity load, including reductions in the performance of acquirers (e.g., Kusewitt, 1985; Laamanen & Keil, 2008; Zorn et al., 2019), changes in their organizational structure (e.g., Barkema & Schijven, 2008), and – as shown in this study – changes in their acquisition behavior. Following this rationale, researchers could, for instance, examine the boundary conditions of these individual responses or even assess whether a hierarchy of firm-level responses to activity load exists. Fourth, scholars could investigate the effect of activity load in non-M&A research contexts, such as strategic alliances (e.g., Shi & Prescott, 2011, 2012), international expansion (e.g., Vermeulen & Barkema, 2002), and organizational change (e.g., Kelly & Amburgey, 1991; Miller & Friesen, 1980). Doing so, future studies would not only explain firm behavior that is relevant to their respective fields but also add to our understanding of the effect of activity load.

6. Conclusion

In this study, I present first empirical evidence on the effect of an acquirer's activity load from acquisitions on that firm's acquisition behavior, with a particular emphasis on how the activity load from acquisitions causes acquirers to deviate from their established acquisition patterns. For this, I exploit changes in the acquisition activity of acquirers in my panel of the 300 largest Fortune Global 500 firms over the 1990-2010 period. My results are based on a hybrid logit model, in which I regress seven operationalizations of an acquirer's acquisition momentum on different definitions of activity load. As predicted, increases in an acquirer's activity load from acquisitions, on average, reduce that firm's ability to maintain its acquisition momentum. Also, my results reveal that acquirers potentially switch from targets in a higher-complexity target firm category to targets in a lower-complexity target firm category to reduce their activity load burden while maintaining their overall momentum. Albeit this finding cautiously indicates that the activity load from acquisitions can explain deviations from established acquisition patterns, the observed effect strongly varies with the operationalization of momentum and, thus, warrants further investigation of this phenomenon. Similarly, I obtain ambiguous results in my analysis of heterogeneity in firm responses arising from differences in the absorptive capacity of firms.

The negative average acquisition response of firms can be explained by an overload situation that is caused by high levels of activity load. That is, as the inertial pressures of momentum induce acquirers to engage in further acquisitions (Amburgey & Miner, 1992; Miller & Friesen, 1980, 1982), the steady increase in acquisition activity directly translates into a higher activity load until the cognitive burden borne by managers exceeds their ability to process this activity load, creating a situation of information overload (Castellaneta & Zollo, 2015). As the strains of information overload are instantly felt by managers and, thus, must be immediately acted on, acquirers are forced to decrease their activity load by reducing their acquisition volume. Yet, while this mechanism is supported by multiple theories, such as Penrosian resource-based logic and the BTF (e.g., Cyert & March, 1963; March & Simon, 1958; Simon, 1945), it is not directly measured in this study and, thus, needs to be further examined to verify its causal interpretation.

Finally, my findings have important implications for scholars and practitioners. That is, on the one hand, this thesis helps scholars better understand a phenomenon of high practical relevance: Deviations from temporal acquisition patterns. Beyond this contribution, scholars benefit from new empirical insights that add to the debate on strategic momentum and past studies of activity load and absorptive capacity. On the other hand, this study makes managers more cognizant of the implications of activity load and provides them with a set of practical actions. Specifically, my findings suggest that managers can actively manage the activity load from acquisitions by (i) regulating the volume of acquisitions, (ii) acquiring targets that are less complex to acquire or integrate, and (iii) increasing the absorptive capacity of their organization. This option space thus suggests that managers are in the driver's seat and ideally should flip the switch before the strains imposed by a firm's activity load do. Therefore, the concept of activity load offers plenty of relevant insights and, thus, should be high up on the research agendas of strategy scholars.

References

Ai, C., & Norton, E. C. (2003). Interaction terms in logit and probit models. *Economics Letters*, *80*(1), 123–129. https://doi.org/10.1016/S0165-1765(03)00032-6

Allison, P. D. (2005). *Fixed Effects Regression Methods for Longitudinal Data Using SAS*. SAS Institute.

Allison, P. D. (2009). *Fixed Effects Regression Models*. SAGE Publications.

Amburgey, T. L., & Miner, A. S. (1992). Strategic Momentum: The Effects of Repetitive, Positional, and Contextual Momentum nn Merger Activity. *Strategic Management Journal*, *13*(5), 335–348.

Audia, P. G., & Greve, H. R. (2006). Less likely to fail: Low performance, firm size, and factory expansion in the shipbuilding industry. *Management Science*, *52*(1), 83–94.

Barkema, H. G., & Schijven, M. (2008). Toward unlocking the full potential of acquisitions: The role of organizational restructuring. *Academy of Management Journal*, *51*(4), 696–722.

Baum, J. A., Li, S. X., & Usher, J. M. (2000). Making the next move: How experiential and vicarious learning shape the locations of chains' acquisitions. *Administrative Science Quarterly*, *45*(4), 766–801.

Baysinger, B., & Hoskisson, R. E. (1989). Diversification Strategy and R&D Intensity in Multiproduct Firms. *Academy of Management Journal*, *32*(2), 310–332.

Beck, N., Bruderl, J., & Woywode, M. (2008). Momentum or deceleration? Theoretical and methodological reflections on the analysis of organizational change. *Academy of Management Journal*, *51*(3), 413–435.

Bertrand, O., Betschinger, M.-A., & Laamanen, T. (2019). Effects of subnational regional corruption on growth strategies in emerging economies: Evidence from Russian domestic and international M&A activity. *Global Strategy Journal*, *9*(2), 303–332.

Biggadike, R. (1979). The Risky Business of Diversification. *Harvard Business Review*, *57*(3), 103–111.

Breen, R., Karlson, K. B., & Holm, A. (2018). Interpreting and Understanding Logits, Probits, and Other Nonlinear Probability Models. *Annual Review of Sociology*, *44*, 39–54.

Bryman, A., & Bell, E. (2011). *Business Research Methods* (3rd ed.). Oxford University Press.

Carow, K., Heron, R., & Saxton, T. (2004). Do early birds get the returns? An empirical investigation of early-mover advantages in acquisitions. *Strategic Management Journal*, *25*(6), 563–585.

Castellaneta, F., & Zollo, M. (2015). The Dimensions of Experiential Learning in the Management of Activity Load. *Organization Science*, *26*(1), 140–157.

Certo, S. T., & Semadeni, M. (2006). Strategy research and panel data: Evidence and implications. *Journal of Management*, *32*(3), 449–471.

Certo, S. T., Withers, M. C., & Semadeni, M. (2017). A tale of two effects: Using longitudinal data to compare within- and between-firm effects. *Strategic Management Journal*, *38*(7), 1536–1556.

Chatterjee, A., & Hambrick, D. C. (2007). It's all about me: Narcissistic chief executive officers and their effects on company strategy and performance. *Administrative Science Quarterly*, *52*(3), 351–386. https://doi.org/10.2189/asqu.52.3.351

Chatterjee, S. (2009). The Keys to Successful Acquisition Programmes. *Long Range Planning*, *42*(2), 137–163.

Collins, J. D., Holcomb, T. R., Certo, S. T., Hitt, M. A., & Lester, R. H. (2009). Learning by doing: Cross-border mergers and acquisitions. *Journal of Business Research*, *62*(12), 1329–1334.

Cools, K., Gell, J., Kengelbach, J., & Roos, A. (2007). The Brave New World of M&A: How to Create Value from Mergers and Acquisitions. *The Boston Consulting Group*. https://www.bcg.com/documents/file15069.pdf

Cyert, R. M., & March, J. G. (1963). *A Behavioral Theory of the Firm*. Prentice-Hall.

Daft, R. L., & Weick, K. E. (1984). Toward a Model of Organizations as Interpretation Systems. *Academy of Management Review*, *9*(2), 284–295.

Dierickx, I., & Cool, K. (1989). Asset stock accumulation and sustainability of competitive advantage. *Management Science*, *35*(12), 1504–1511. https://doi.org/10.1287/mnsc.35.12.1504

Dobbins, L., & Jacob, M. (2016). Do corporate tax cuts increase investments? *Accounting and Business Research*, *46*(7), 731–759.

Ellis, K. M., Reus, T. H., Lamont, B. T., & Ranft, A. L. (2011). Transfer effects in large acquisitions: How size-specific experience matters. *Academy of Management Journal*, *54*(6), 1261–1276.

Finkelstein, S., & Haleblian, J. J. (2002). Understanding acquisition performance: The role of transfer effects. *Organization Science*, *13*(1), 36–47.

Gamache, D. L., McNamara, G., Mannor, M. J., & Johnson, R. E. (2015). Motivated to acquire? The impact of CEO regulatory focus on firm acquisitions. *Academy of Management Journal*, *58*(4), 1261–1282.

Greve, H. R. (2002). Sticky aspirations: Organizational time perspective and competitiveness. *Organization Science*, *13*(1), 1–17.

Haleblian, J. J., & Finkelstein, S. (1999). The influence of organizational acquisition experience on acquisition performance: A behavioral learning perspective. *Administrative Science Quarterly*, *44*(1), 29–56.

Haleblian, J. J., Kim, J. Y. J., & Rajagopalan, N. (2006). The influence of acquisition experience and performance on acquisition behavior: Evidence from the US commercial banking industry. *Academy of Management Journal*, *49*(2), 357–370.

Haleblian, J. J., Pfarrer, M. D., & Kiley, J. T. (2017). High-Reputation Firms and Their Differential Acquisition Behaviors. *Strategic Management Journal*, *38*(11), 2237–2254.

Haspeslagh, P. C., & Jemison, D. B. (1991). *Managing Acquisitions: Creating Value through Corporate Renewal*. Free Press.

Haunschild, P. R. (1993). Interorganizational Imitation: The Impact of Interlocks on Corporate Acquisition Activity. *Administrative Science Quarterly*, *38*(4), 564–592. https://doi.org/10.2307/2393433

Haunschild, P. R., & Beckman, C. M. (1998). When do interlocks matter?: Alternate sources of information and interlock influence. *Administrative Science Quarterly*, *43*(4), 815–844.

Hayward, M. L. A. (2002). When do firms learn from their acquisition experience? Evidence from 1990-1995. *Strategic Management Journal*, *23*(1), 21–39. https://doi.org/10.1002/smj.207

Herold, F. (2019). Personal Taxes and Corporate Investment. *Junior Management Science*, *4*(1), 81–100. https://doi.org/10.5282/jums/v4i1pp81-100

Hoetker, G. (2007). The use of logit and probit models in strategic management research: Critical issues. *Strategic Management Journal*, *28*(4), 331–343.

Iyer, D. N., & Miller, D. (2008). Performance feedback, slack, and the timing of acquisitions. *Academy of Management Journal*, *51*(4), 808–822.

Judge, G. G., Hill, R. C., Griffiths, W., Lutkepohl, H., & Lee, T.-C. (1982). *Introduction to the theory and practice of econometrics*. Wiley.

Kalnins, A. (2018). Multicollinearity: How common factors cause Type 1 errors in multivariate regression. *Strategic Management Journal*, *39*(8), 2362–2385.

Karaevli, A. (2007). Performance consequences of new CEO outsiderness': Moderating effects of pre- and post-succession contexts. *Strategic Management Journal*, *28*(7), 681–706.

Karim, S., & Mitchell, W. (2000). Path-dependent and path-breaking change: Reconfiguring business resources following acquisitions in the US medical sector, 1978-1995. *Strategic Management Journal*, *21*(10-11), 1061–1081.

Karlson, K. B., Holm, A., & Breen, R. (2012). Comparing Regression Coefficients Between Same-sample Nested Models Using Logit and Probit: A New Method. *Sociological methodology*, *42*(1), 286–313. https://doi.org/10.1177/0081175012444861

Kelly, D., & Amburgey, T. L. (1991). Organizational Inertia and Momentum: A Dynamic Model of Strategic Change. *Academy of Management Journal*, *34*(3), 591–612.

Kengelbach, J., & Roos, A. (2011). *Riding the Next Wave in M&A: What are the Opportunities to Create Value?* The Boston Consulting Group. Retrieved August 15, 2021, from https://www.bcg.com/publications/2011/riding-the-next-wave-in-m-and-a

Kim, J.-Y., Finkelstein, S., & Haleblian, J. J. (2015). All Aspirations Are Not Created Equal: The Differential Effects of Historical and Social Aspirations on Acquisition Behavior. *Academy of Management Journal*, *58*(5), 1361–1388.

Kim, J.-Y., Haleblian, J. J., & Finkelstein, S. (2011). When Firms Are Desperate to Grow via Acquisition: The Effect of Growth Patterns and Acquisition Experience on Acquisition Premiums. *Administrative*

Science Quarterly, *56*(1), 26–60. https://doi.org/10.2189/asqu.2011.56.1.026

King, D. R., Dalton, D. R., Daily, C. M., & Covin, J. G. (2004). Meta-analyses of post-acquisition performance: Indications of unidentified moderators. *Strategic Management Journal*, *25*(2), 187–200.

Klarner, P., & Raisch, S. (2013). Move to the Beat – Rhythms of Change and Firm Performance. *Academy of Management Journal*, *56*(1), 160–184.

Kusewitt, J. B. (1985). An Exploratory Study of Strategic Acquisition Factors Relating to Performance. *Strategic Management Journal*, *6*(2), 151–169.

Laamanen, T., & Keil, T. (2008). Performance of serial acquirers: Toward an acquisition program perspective. *Strategic Management Journal*, *29*(6), 663–672.

Langley, A., & Abdallah, C. (2011). Templates and Turns in Qualitative Studies of Strategy and Management. In *Building Methodological Bridges*. Emerald Group Publishing Ltd.

Lasserre, P. (2003). *Global Strategic Management (1st ed.)* Macmillan International Higher Education.

Lee, G. K., & Lieberman, M. B. (2010). Acquisition vs. internal development as modes of market entry. *Strategic Management Journal*, *31*(2), 140–158. https://doi.org/10.1002/smj.804

Levinthal, D., & March, J. G. (1981). A Model of Adaptive Organizational Search. *Journal of Economic Behavior & Organization*, *2*(4), 307–333.

Levitt, B., & March, J. G. (1988). Organizational Learning. *Annual Review of Sociology*, *14*, 319–340. https://doi.org/10.1146/annurev.so.14.080188.001535

Malmendier, U., & Tate, G. (2005). CEO overconfidence and corporate investment. *Journal of Finance*, *60*(6), 2661–2700.

Malmendier, U., & Tate, G. (2008). Who makes acquisitions? CEO overconfidence and the market's reaction. *Journal of Financial Economics*, *89*(1), 20–43. https://doi.org/10.1016/j.jfineco.2007.07.002

March, J. G., & Simon, H. A. (1958). *Organizations*. Wiley.

McNamara, G. M., Haleblian, J. J., & Dykes, B. J. (2008). The performance implications of participating in an acquisition wave: Early mover advantages, bandwagon effects, and the moderating influence of industry characteristics and acquirer tactics. *Academy of Management Journal*, *51*(1), 113–130.

Miller, D., & Friesen, P. H. (1980). Momentum and Revolution in Organizational Adaptation. *Academy of Management Journal*, *23*(4), 591–614.

Miller, D., & Friesen, P. H. (1982). Innovation in Conservative and Entrepreneurial Firms: Two Models of Strategic Momentum. *Strategic Management Journal*, *3*(1), 1–25. https://doi.org/10.1002/smj.4250030102

Mize, T. D. (2019). Best Practices for Estimating, Interpreting, and Presenting Nonlinear Interaction Effects. *Sociological Science*, *6*, 81–117. https://doi.org/10.15195/v6.a4

Mize, T. D., Doan, L., & Long, J. S. (2019). A General Framework for Comparing Predictions and Marginal Effects Across Models. *Sociological Methodology*, *49*(1), 152–189.

Nelson, R., & Winter, S. G. (1982). *The Evolutionary Theory of Economic Change*. Harvard University Press.

Neter, J., Wasserman, W., & Kutner, M. H. (1989). *Applied Linear Regression Models (2nd ed.)* Irwin.

Ocasio, W. (1997). Towards an attention-based view of the firm. *Strategic Management Journal*, *18*(S1), 187–206.

Palepu, K. (1985). Diversification Strategy, Profit Performance and The Entropy Measure. *Strategic Management Journal*, *6*(3), 239–255. https://doi.org/10.1002/smj.4250060305

Penrose, E. (1959). *The Theory of the Growth of the Firm (4th ed.)* Oxford University Press.

Prahalad, C. K., & Bettis, R. A. (1986). The Dominant Logic: A New Linkage Between Diversity and Performance. *Strategic Management Journal*, *7*(6), 485–501. https://doi.org/10.1002/smj.4250070602

Refinitiv. (n.d.). Mergers and Acquisitions from Refinitiv. Retrieved August 15, 2021, from https://www.refinitiv.com/content/dam/marketing/en_us/documents/fact-sheets/mergers-and-acquisitions-fact-sheet.pdf

Ronen, S., & Shenkar, O. (2013). Mapping world cultures: Cluster formation, sources and implications. *Journal of International Business Studies*, *44*(9), 867–897.

Roumeliotis, G., & Barbaglia, P. (2019). *Dealmakers eye cross-border M&A recovery as mega mergers roll on*. Retrieved August 15, 2021, from https://www.reuters.com/article/us-global-deals-idUSKBN1YZ0YZ

Sanders, W. G. (2001). Behavioral responses of CEOs to stock ownership and stock option pay. *Academy of Management Journal*, *44*(3), 477–492.

Schipper, K., & Thompson, R. (1983). Evidence on the Capitalized Value of Merger Activity for Acquiring Firms. *Journal of Financial Economics*, *11*(1-4), 85–119.

Schunck, R. (2013). Within and between estimates in random-effects models: Advantages and drawbacks of correlated random effects and hybrid models. *Stata Journal*, *13*(1), 65–76.

Schunck, R., & Perales, F. (2017). Within- and between-cluster effects in generalized linear mixed models: A discussion of approaches and the xthybrid command. *Stata Journal*, *17*(1), 89–115.

Seo, J., Gamache, D. L., Devers, C. E., & Carpenter, M. A. (2015). The role of CEO relative standing in acquisition behavior and CEO pay. *Strategic Management Journal*, *36*(12), 1877–1894.

Shaver, J. M. (2006). A paradox of synergy: Contagion and capacity effects in mergers and acquisitions. *Academy of Management Review*, *31*(4), 962–976.

Shi, W., Hoskisson, R. E., & Zhang, Y. A. (2017). Independent Director Death and CEO Acquisitiveness: Build an Empire or Pursue a Quiet Life? *Strategic Management Journal*, *38*(3), 780–792.

Shi, W., & Prescott, J. E. (2011). Sequence Patterns of Firms' Acquisition and Alliance Behaviour and their Performance Implications. *Journal of Management Studies*, *48*(5), 1044–1070.

Shi, W., & Prescott, J. E. (2012). Rhythm and Entrainment of Acquisition and Alliance Initiatives and Firm Performance: A Temporal Perspective. *Organization Studies*, *33*(10), 1281–1310.

Shi, W., Sun, J., & Prescott, J. E. (2012). A Temporal Perspective of Merger and Acquisition and Strategic Alliance Initiatives: Review and Future Direction. *Journal of Management*, *38*(1), 164–209.

Shinkle, G. A. (2012). Organizational Aspirations, Reference Points, and Goals: Building on the Past and Aiming for the Future. *Journal of Management*, *38*(1), 415–455.

Simon, H. A. (1945). *Administrative Behavior*. Macmillan.

Train, K. E. (1986). *Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand (1st ed.)* MIT Press.

Train, K. E. (2009). *Discrete Choice Methods with Simulation (2nd ed.)* Cambridge University Press.

Vermeulen, F., & Barkema, H. (2001). Learning through acquisitions. *Academy of Management Journal*, *44*(3), 457–476.

Vermeulen, F., & Barkema, H. (2002). Pace, rhythm, and scope: Process dependence in building a profitable multinational corporation. *Strategic Management Journal*, *23*(7), 637–653.

Wolfolds, S. E., & Siegel, J. (2019). Misaccounting for endogeneity: The peril of relying on the Heckman two-step method without a valid instrument. *Strategic Management Journal*, *40*(3), 432–462.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data (2nd ed.)* MIT Press.

Zajac, E. J. (1990). CEO Selection, Succession, Compensation and Firm Performance: A Theoretical Integration and Empirical Analysis. *Strategic Management Journal*, *11*(3), 217–230. https://doi.org/10.1002/smj.4250110304

Zorn, M. L., Sexton, J., Bhussar, M. S., & Lamont, B. T. (2019). Unfinished Business: Nested Acquisitions, Managerial Capacity, and Firm Performance. *Journal of Management*, *45*(4), 1488–1516.

Junior Management Science

www.jums.academy
ISSN: 2942-1861

Decarbonizing Travel Decisions by Using Digital Nudges

Patrick Riedlsperger

Technical University of Munich

Abstract

The current climate crisis was caused by our everyday, individual decision-making. People have the opportunity to decide between options that contain more or less greenhouse gases. This is particularly relevant for the travel industry which has historically been a major contributor to global emissions. The nudging concept introduced by Sunstein and Thaler (2021) can help people enhance their decision-making to promote environmental stewardship. Every consumption decision in travel is an opportunity as it can be 'decarbonized' to a greener outcome. This thesis provides evidence that the intervention technique is effective to lead to more sustainable decision-making in a digital travel booking process. This research project used a simulated booking process to compare the effectiveness of different digital nudges. Users could choose different options in their booking in the realm of transport, accommodation and restaurants. Overall, 456 online participants completed the process. The digital experiment used one regular booking process, which was used as a reference group, and 9 different types of digital nudges. The effectiveness of the nudges was analyzed by using a binary logistic regression model. Of the 9 experiments which included digital nudging interventions, 6 produced statistically significant results. The most effective nudge in the experiment used a social norm intervention. After its application to the process, odds were more than 4 times higher that users chose the most sustainable option that contained the least amount of greenhouse gases. In general, all regression coefficients (B) were positive, with odds ratios Exp(B) between 2.471 and 4.419. The results of this thesis support the view that nudges are an effective tool to drive more sustainable behavior. The results showed that digital nudges led to the booking of the most sustainable travel offers. User interface designers and other choice architects can use the findings of this thesis to reduce greenhouse gas emissions in travel as one of the many steps we must undertake to fight global warming and its drastic impacts on our economy and society.

Keywords: choice intervention; digital nudging; nudge theory; sustainability; travel

1. Introduction

This thesis addresses one of the most important challenges our society faces in this century. While writing it, Fountain (2022) reported for the New York Times that heat waves in Europe are increasing in frequency and intensity at a faster rate than almost any other part of the planet. Temperature highs hit new records across the continent this summer. Rising temperatures, drought, frequent wildfires, shifting rainfall patterns, melting glaciers and the rise of the average global sea level prove that the impacts of climate change are underway. To mitigate climate change, we must reduce or at least prevent emissions linked to human activities (European Environment Agency, 2021). Schellnhuber (2021) is the originator of the 2 degrees Celsius global temperature target and described the climate crisis as similar to the COVID-19 crisis - just substantially bigger by dimensions and consequences. We have acute emergencies, loss of hundreds of thousands of lives, and other socioeconomic impacts that widen the wealth gap. According to the renowned researcher, the potential damage that climate change can cause is even greater than COVID-19 by a factor of a hundred or even a thousand. In both cases, however, it is a question of acting in a timely manner. When it comes to climate change, we need to turn things around in the next three decades or we will reach an irreversible state.

Global warming is caused by humans and their emissions

DOI: https://doi.org/10.5282/jums/v9i1pp1178-1210

of greenhouse gases, as they blanket the Earth and trap the sun's heat (United Nations, 2022). Every day, each of us makes a host of purchase and consumption decisions. Embedded in every decision is some amount of greenhouse gas. This means all of our many small, individual actions add up to a substantial impact on the climate. The way forward will be determined by the decisions made by millions and our ability to pivot many of those decisions toward a greener outcome (Amram & Kulatilaka, 2009). People can reduce their footprint by buying products such as the Impossible Burger or electric cars. They can change their own personal consumption habits, so that clean products can achieve scale and their costs go down. Decarbonization will be the most incredible feat achieved by humankind (Gates, 2021).

Besides strict regulatory measures and the potential to reduce carbon footprints when delivering services or making products, we must find complementary ways to react to the climate crisis at an individual level of decision-making. During my Executive MBA program at TU Munich and HSG St. Gallen, I came across the book "Nudge: Improving Decisions About Health, Wealth, and Happiness" by Sunstein and Thaler (2021). Nudging is a theory in the field of behavioral economics. In short, a Nudge aims to lead consumers to make more efficient decisions for themselves, and as a result, for society. It challenges the idea that rational humans always choose the option that is 'best for them' and unpacks how to 'nudge' individuals into making better decisions (Spiliakos, 2017).

In their bestselling book, Sunstein and Thaler (2021, p. 8) also mention the potential to use nudging interventions to promote more sustainable consumption behaviors.

> *"Private companies that want to make money and to do good can benefit by creating environmentally friendly nudges, helping to reduce air pollution and the emission of greenhouse-gases."*

Accelerating climate action by reducing our emissions is of utmost importance for travel. This industry is highly vulnerable to climate change. At the same time, it contributes to some of the highest emissions of greenhouse gases of any industry worldwide. It must comply with the global COP26 commitment to halve our emissions by 2030 and achieve net zero by 2050. There is a growing consensus among stakeholders in the travel industry that the future and resilience of tourism will depend on the ability to cut emissions by 50% by 2030 (UNWTO, 2019).

My motivation for this research project was to seize and further explore the idea of Sunstein and Thaler to use nudging interventions for the reduction of greenhouse gases. Based on the relevance of climate action to travel, my goal was to explore how businesses can use Nudging to reduce emissions and fight our climate crisis. The central actor in climate change and this thesis is the individual consumption decisions we make every day and the related impacts.

Nudging seems to be an attractive opportunity to fight climate change, without jeopardizing freedom of choice. Many researchers, including Sunstein (2019), emphasize the importance of evidence when implementing nudges, as some interventions seem promising in the abstract, but fail in practice. Sunstein (2017b) even published an article on "Nudges that fail" explaining why nudging is sometimes ineffective, or at least less effective than we hope and expect. Empirical tests, including randomized controlled trials, are the solution to overcome this issue. Researchers should investigate which types of nudges tend to have larger effects on outcomes. Empirical tests can reveal the best nudging techniques to achieve a specific goal.

Schneider et al. (2018) encourages fellow scholars to engage in research on nudging in digital choice environments, because of the ubiquitous digitalization of our private and professional lives. According to Lehner et al. (2016), the evidence base for the effectiveness of nudges in sustainable consumption remains an important research topic. Furthermore, to take full advantage of digital nudging, Mirsch et al. (2018) stress that interventions must be developed systematically, applied on the user, and then tested for their effectiveness. Quantitative research to test the effectiveness of Nudging has been conducted before. Most research in this domain is limited to a certain context such as energy (Allcott, 2011; Ebeling & Berger, 2015; Lade et al., 2020), diet (Hanks et al., 2012; Jesse et al., 2021; Tett, 2021), or finance (Franklin et al., 2019; García & Vila, 2020; Thaler & Benartzi, 2004). In a travel-specific context current research projects on nudging people towards more sustainable options is restricted to transport (Nijhuis, 2020) or the qualitative feedback of research participants (Andersson, 2019).

Furthermore, systematic literature reviews across domains (Hummel & Maedche, 2019; Mertens et al., 2022; Szaszi et al., 2018) investigate whether nudging interventions across techniques and behavioral domains are effective. If these results can be generalized is questionable as the virtue of nudges is context specific (Kosters & van der Heijden, 2015; Sunstein & Thaler, 2021)

The purpose of this thesis is to reduce greenhouse gas emissions. More specifically, this study aims to quantitatively identify the most effective ways to nudge users towards more sustainable options in a travel-specific context. The goal is to add value to the topic by testing if nudges are effective for more sustainable decisions and which nudges are most effective. Upon the completion of this research project, the following questions should be answered:

> *Are Nudges an effective way to promote more sustainable travel decisions?*
>
> *Which Nudges are effective in promoting more sustainable travel decisions?*

Companies in the travel industry should be able to utilize the results to drive sales of the most sustainable products on their channels. It is one of the many small steps we need to take for this major challenge of humankind to reach net zero by 2050 and tackle climate change.

2. Theoretic Background

2.1. The Climate change challenge

In 'An Inconvenient Truth', Gore (2006) made an impassioned call for immediate climate action. The film focused on the grave state of our environment, earned two Academy Awards and was one of the highest grossing documentaries of all time. A major audience, for the first time, was confronted with one of the biggest challenges of our modern society. The film changed viewers' attitudes towards global warming. 73% of all viewers even indicated that they would change their habits because of the climate crisis (Mazar et al., 2020; Nielsen, 2007).

However, beliefs and concerns often do not result in climate action. Hornsey et al. (2016) meta-analyzed 196 studies and polls, and found that environmental-friendly attitudes and intentions can only be modestly associated with environmental-friendly behavior. Jacobsen (2011) tested if people took action after they watched 'An Inconvenient Truth' and found that effects faded quickly after initial actions. It seems that we as a society fail to act to protect the environment, because we lack concern on the issue. Our concern about climate change has grown globally, but on an individual level, our decision behaviors do not seem to reflect that concern (Mazar et al., 2020).

The primary driver of the human-caused climate change is the rise of atmospheric levels of carbon dioxide and other greenhouse gases. Since the Industrial Revolution, humans have released nearly 2,500 metric gigatons of CO2 into earth's atmosphere. If we do not implement any significant changes, global temperatures could increase by 2.3 degrees Celsius by 2050. Multiple scientists predict that this could be a point of no return. Feedback loops such as the thawing of permafrost, which will lead to an additional emission of greenhouse gases could ultimately turn earth into a 'hothouse' state. Potential impacts will rise over time if levels of greenhouse gases in our atmosphere continue to rise. Corporates and governments alike must integrate climate change into their decision-making to accelerate the pace and scale of adaptation, and decarbonize at scale to mitigate risks (Woetzel et al., 2020).

Between 2022 and 2026, the annual mean global surface temperature is predicted to be 1.1 and 1.7 degrees Celsius higher than preindustrial levels. The chance of this five-year mean being higher than the last five years (2017-2021) is 93% (World Meteorological Organization, 2022). Even if our emissions came to a sudden halt, Earth's atmosphere will continue to warm. This illustrates the difficulty of reversing climate change (Frölicher et al., 2013).

The amount of warming largely depends on the choices we make now and in the next decades. The IPCC (2021) illustrates five different scenarios, based on Shared Socioeconomic Pathways (SSPs). These scenarios include natural events like volcano activity and a broad range of social and economic forces, which are driving greenhouse gases. SSP1-1.9 represents the low end of emissions, leading to a warming below 1.5 degrees Celsius in 2100. Together with the SSP1-2.6 scenario these calculations are based on declining CO2 emissions to net zero around 2050. At the other end, the SSP5-8.5 scenario calculates that humans will double their emissions by 2100 compared to today's levels. The report even reaffirms a near-linear relationship between cumulative CO2 emission and the global warming they cause. Each 1,000 gigatons is assessed to cause an increase of global surface temperature of 0.45 degrees Celsius.

The global momentum towards decarbonization continues to grow. Most developed countries, leading companies and other organizations have reached broad consensus to pursue net-zero emissions (Engel et al., 2022). The international, legally-binding "Paris Agreement" (2016) is the most far-reaching in our history. Currently, 194 states and the European Union signed to:

> *Article 2 – 1. (a) Holding the increase in the global average temperature to well below 2°C above pre-industrial levels and pursuing efforts to limit the temperature increase to 1.5°C above pre-industrial levels, recognizing that this would significantly reduce the risks and impacts of climate change;*

According to climate scientists many of the direst effects of human-caused climate change can still be avoided. There are severe and currently uncertain impacts such as ice-sheet collapse, deforestation or an abrupt change in ocean circulation. However, the biggest uncertainty in all climate change projections is how humans will act. It is still possible to limit global warming to within 1.5 degrees Celsius by immediate, rapid and large-scale reduction of all greenhouse gases. The climate future can be changed, if we change our behavior with new ideas and actions (Tollefson, 2021). If we are able to transform our economy, reach political agreements and public buy-in to sharply reduce our emissions, there is still hope to limit the destruction caused by the climate crisis (Fischetti, 2021). Environmental improvements in companies were traditionally focused on pollution control. However, companies and regulators must find ways to prevent environmental harmful emissions before they occur (Porter & van der Linde, 1995). To contribute to systemic change and substantial reduction of our greenhouse gas accumulation, we must also empower individuals. Climate change is the aggregation of billions of individual decisions. Climate actions such as living car-free, avoiding airplane travel or switching to a plant-based diet have tremendous potential to reduce the pace of greenhouse gas accumulation in our atmosphere (Wynes & Nicholas, 2017).

2.2. The travel industry and it's footprint

The World Travel & Tourism Council (2022) reports on the economic and employment impact of the travel industry for 185 countries around the world. According to the report, the travel industry accounted for 1 in 4 of all newly-created jobs across the world in 2019 when accounting for its direct, indirect and induced impacts. The sector contributed 10.3%

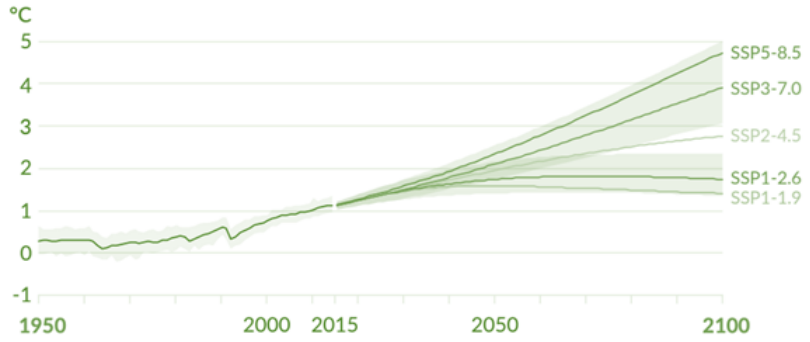


Figure 1: *Global surface temperature change relative to 1850-1900 (IPCC, 2021)*

to global GDP and comprised 6.8% of total exports. Due to the necessary restrictions in mobility during the COVID-19 pandemic, global GDP share decreased to 5.3% in 2020 and recovered to 6.1% in 2021.

Research undertaken by Lenzen et al. (2018) quantified tourism-related global carbon flows between 160 countries and indicated that the travel industry will constitute a growing part of our greenhouse gas emissions. About 8% of global greenhouse gas emissions are currently tied to tourism. The global footprint of tourism increased from 3.9 to 4.5 metric gigatons carbon dioxide between 2009 and 2013. Consumer demand for travel has grown much faster than their consumption of other products and services. This global demand is outstripping the decarbonization efforts of tourism operations. When we evaluate the carbon burden of different travel activities, there is significant variation, with aviation being the most critical component. The majority of this footprint is caused by high-income countries. Breaking down carbon emissions to different tourism-related activities, the highest proportions occur in transport (especially by air), goods (shopping) and hospitality (accommodation and restaurants).

The number of international travelers is expected to reach 1.8 billion per year in 2030. Based on a current scenario, transport related CO2 emissions alone from travelers will grow 25% from 2016 levels. The predicted growth will bring opportunities such as socioeconomic development and job creation, but also challenges to meet climate targets (UNWTO, 2019).

According to the World Tourism Organization (2008) consumers should be encouraged to consider the climate and environmental impacts of their options before making a decision. Whenever possible, tourists should try to reduce their carbon footprint and opt for environmentally friendly activities at the destination. Exemplary measures are raising awareness for the issue among customers, promoting public modes of transport, improving awareness and transparency around emissions, and creating standardized carbon footprint labeling on all tourism products.

2.3. Our operating systems

People employ a very limited number of heuristics to simpler judgmental operations. While this reduction in complexity of our decision-making processes is usually effective, it can also lead to severe and systematic errors (Tversky & Kahneman, 1974). In his behavioral economics memoir, *"Thinking Fast and Slow",* Kahneman (2013) says that our minds process information in two distinct ways. In System 2, our minds are concerned with effortful mental activities that demand brainpower, including complex computations. However, most of the time, people operate on System 1 which allocates attention to automated, intuitive decisions with little or no effort and no sense of voluntary control. System 1 is the dominating mode, which is also guiding and steering our analytical System 2 to a very large extent.

The theory of the Homo Economicus is fiction. Real humans embrace irrelevant information, see patterns where none exist and are subject to serious inertia. All our minds are dichotomous. The first half's seize is resolute, farsighted and reflective. However, the other reptile half often seizes the levers of choice in an impulsive, myopic and emotional way. This is why we smoke, drink, eat too much, or exercise and save too little. People can better be described as Homer Economicus than the theoretical rational ideal. The key claim in behavioral economics is not that people are fallible; it is that humans make mistakes systematically (Leonard, 2008).

2.4. Stone Time Psychological Biases

Human social behavior has developed in the course of the evolution. However, people still have behavioral patterns, which stem from the Stone Age and can lead us to erroneous decision-making. Our modern environmental problems are caused by these biases. At the same time, we can employ identical biases, to systematically develop influence strategies towards environmental conservation and change (Thorun et al., 2017; Vugt et al., 2014).

New research from scholars at INSEAD and the University of Southern California has shown that there is an attitude-behavior gap in sustainable habits. Governments and businesses can reduce this gap by interventions that draw on the insights of psychology and behavioral economics (Mazar et al., 2020). People make decisions quickly under pressure,

Table 1: *Economic impacts of the travel industry from 2019 to 2021 – Adapted from* World Travel & Tourism Council (2022)

	Total GDP contribution (year-on-year change)		*Total jobs in the travel industry*	
2019	9,630 billion USD	+4.7%	333 million	1 in 10 jobs
2020	4.775 billion USD	-50.4%	271 million	1 in 12 jobs
2021	5.812 billion USD	+21.7%	289 million	1 in 11 jobs

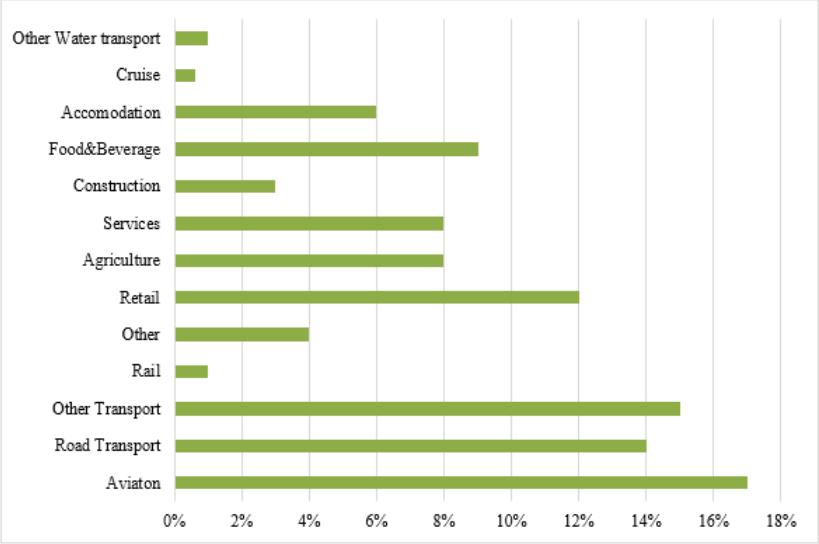


Figure 2: *Carbon Footprint of Global Tourism – Adapted from* World Travel & Tourism Council (2021)

Figure 3: *Homer deciding on System 1 mode, Spock operating in System 2 logic* Gilbert (2021)

based on system 1 and guided by biases and psychological fallacies. There are many pitfalls to reasonable decision making – taking the most beneficial choice (Sperling & Güntner, 2017).

2.5. The Nudging Concept

Nudging uses biases and other systemic errors in our decision-making processes. Instead of appeals, tax incentives or bans, nudging applies psychological methods (Rauner, 2015). Introduced by Harvard professor Cass Sunstein and University of Chicago professor Richard Thaler, these interventions are aimed at getting people to act in their own best interest. Nudges alter people's decision behavior in predicable ways and help them to improve their lives, while maintaining freedom of choice. Hansen (2016, p. 168) defines nudging as an attempt to influence the behavior of people in a predictable way without forbidding or adding any rationally relevant choice options or changing incentives. The goal of nudging is to make life simpler and easier to navigate for choosers. A good example of an application of nudging is a GPS system, which people use in their cars or on their smartphones. The GPS nudges people to steer in a certain direction, while having the freedom of selecting their own route instead (Sunstein, 2019). To count as a nudge, interventions must be easy and cheap to avoid. Nudges are never mandating. Imagine a parent that wants their children to eat healthier. The overall goal is to teach children to decide in their own best interest, which practically means increasing their consumption of fresh fruits and vegetables

Table 2: *Constraints and Obstacles of Psychological Biases – Adapted from* Vugt et al. (2014)

Psychological Bias	**Constraints on behavior change**	**Ways of intervention**
Self-interest	People prioritize personal over collective interests	Persuade individuals to value the collective more than their own interests
Shortsighted	People value the present more than the future	Persuade individuals to value the future more than the present
Status	People value relative over absolute status	Persuade individuals to accept a lower relative status associated with environmental conservation
Social imitation	People copy what others around them are doing	Persuade people to behave environmentally despite not many others behaving in this way
Sensory mechanisms	People ignore threats and dangers they cannot see, smell or touch	Persuade individuals to be concerned about distant, global, and slow-moving environmental problems

and decreasing their consumption of junk food:

- **Nudge:** Fruits and vegetables are placed closer to locations where children usually play to steer them towards healthier choices. At the same time, place junk food is placed in a cupboard, to reduce their consumption of unhealthy foods
- **Not a nudge:** Unhealthy food choices are completely banned from the children's diet and they are forced to only eat fruits and vegetables (Pereira, 2019).

Nudges have been used by both large and small, public and private sector organizations around the world. Most recently, health organizations have used nudges to educate citizens on COVID-19 testing and vaccination. Nevertheless, the main goal of nudges may not always be to protect the chooser. Nudges also help to protect third parties such as our climate (Thaler & Sunstein, 2021). Private companies that want to do good and make money at the same time can implement nudges to reduce air pollution and the emissions (Sunstein & Thaler, 2021).

Nudges are always based on an underlying choice architecture. Similar to traditional architecture, it is crucial to understand that there is no such thing as a neutral design. Small details can have a major impact on people's actual behavior. Good architects are aware that although there is no such thing as a perfect building, they make some choices which will have beneficial effects. For example, workplace interaction may be influenced by the location of the coffee machine when they design an office building. Choice architects have the responsibility of organizing the context in which humans make decisions and have the power to steer people's choices in a direction that will improve their lives. Most people are actually choice architects without realizing it. Some examples of day-to-day choice architects in our lives include:

- A medical doctor describing different medical treatment options to patients,
- A sales manager presenting different products to clients
- A caterer deciding how food and beverages are presented in a cafeteria,
- A web developer who designs interactions on websites (Sunstein & Thaler, 2021).

2.6. Why Libertarian Paternalism

Nudging is a technique that uses the idea of Libertarian Paternalism. This term might not be endearing to many readers as both concepts seem to be contradictory (Sunstein & Thaler, 2021). As many economists are libertarians, the term paternalism may even be derogatory. This is based on the false assumption that people always make choices which are in their best interest. While people should be "free to choose" as Friedman and Friedman (1990) put it, paternalistic measures help them to take better decisions for themselves. Furthermore, in many cases, a choice architect must make choices affecting others and paternalism does not always involve coercion by definition. As in the caterer example, the choice architect must make a decision on how to present food and beverages in a cafeteria. Libertarian Paternalism means that the person can arrange the products in an order, which benefits the health of guests. However, there is the liberty of every individual to make their own selection (Sunstein & Thaler, 2003). Nudges can alter behaviors towards climate-friendly actions and are the subject of enthusiasm to steer without having to resort to 'hard' public regulation (Siipi & Koi, 2022).

2.7. Nudging Toolbox

Choice architecture can succeed in many managerial settings. There are a variety of nudges which have been studied by scholars in many different academic disciplines (Beshears & Kosowsky, 2020). There are different ways to categorize nudges. One example is by educational or non-educational

Table 3: *Categories and examples of policy interventions – Adapted from* House of Lords (2011)

Regulation of the individual	Fiscal measures directed at the individual	Non-regulatory and non-fiscal measures with relation to the individual
	Guide and enable choice	
Eliminate choice (prohibiting goods/services) **Restrict choice** (outlaw smoking in public)	**Fiscal incentives** (tax cuts/breaks) **Fiscal disincentives** (taxation on cigarettes)	**Non-fiscal incentives and disincentives** (time-off work to volunteer) **Persuasion** (marketing campaigns) **Choice Architecture** (nudges)

nudges, which evaluates whether people like or dislike being nudged in a particular way (Sunstein, 2017a), whether it triggers our described system 1 or system 2 (Sunstein, 2016) and how the nudge influences the choice of individuals (Lehner et al., 2016). Sunstein (2019) also mentions there is an exceedingly wide range of interventions and their number and variety is constantly growing. Hence, he created an overview of 10 important nudges, which choice architects can use.

2.7.1. Default rules

Providing a default option is the simplest example of a successful nudge. It is simply what happens if the chooser does nothing. Many people just go with the flow sometimes knowingly and sometimes unknowingly. Default rules are an extremely powerful tool for choice architects to implement. One example of a default rule is in the area of pension policy in the US. In many 401(k) plans, the default is not to join. If you want to join, there is the duty of filling out paperwork. Companies which decided the opposite default option increased enrollment to the pension policy greatly (Thaler, 2009). Defaults can be seen as manufacturer recommendations. They have the potential to enhance customer experience and drive sales. A large national railroad in Europe made a small change to their website where a ticket purchase automatically included a fee-based seat reservation. Before this change was made, only 9% of users chose the reservation option, which increased to 47% after the implementation of the nudge. The railroad earned an additional US$40 million, with only a small fixed cost in programming and infrastructure (Heitmann et al., 2008). Default options can also manage our transition into a carbon-free economy. Default engines of new cars could be set to hybrid or fully electric. Standard temperatures of washing machines could be low and users would need to actively switch to higher temperatures (Berger, 2015). Along with a nationwide energy supplier in Germany, Ebeling and Berger (2015) attempted to use default rules to nudge existing customers to a new green energy contract that stemmed entirely from renewable resources. Setting the default choice to the more sustainable option nearly ten folded purchases of the green energy plan.

2.7.2. Simplification

People struggle to make choices, especially for complex products. The complexity of the information provided greatly affects the outcomes of decisions. Simplification nudges build on the insight that the amount and accessibility of information provided are not the only things that matter to people. Simplifications nudges can support choosers by making information more straightforward and presented in a way that best fits their information processing capabilities and decision-making process. One example of simplifications is food labels. They are often focused on counteracting lifestyle-related health problems such as obesity or diabetes (Mont et al., 2014). Another case of simplification is by labeling the Energy Star brand by the Department of Energy and the Environmental Protection. This label identifies products that meet certain energy efficiency standards. The label increases simplification by decreasing the amount of information that individuals have to process. It allows customers to choose energy efficient options with less research effort (Cooper, 2017).

2.7.3. Use of Social Norms

People tend to make choices based on social influence. Social Norm nudges inform people what other choosers are doing and thereby induce them to alter the same decision (Nahmias, 2019). Humans are nudged by other humans because they tend to think others have better information and understanding of a topic or because they just like to conform to a group. Asch (1951) conducted a series of experiments on how we tend to follow the herd. When the participants were asked to decide individually, and without judgment of

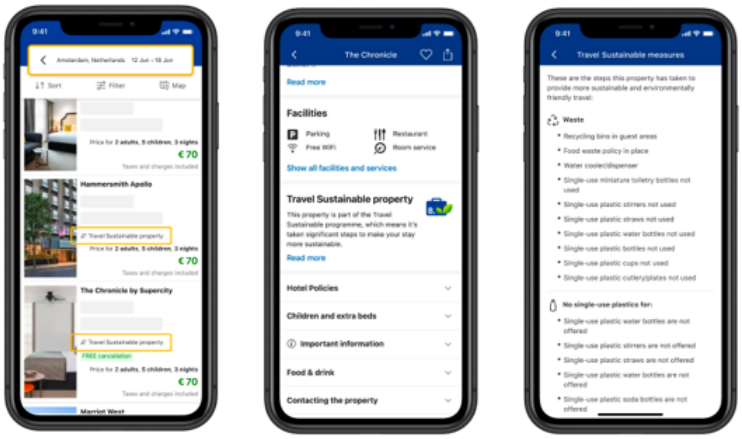


Figure 4: *Booking.com's sustainable travel label – Adapted from Booking.com*

others on a very easy task, they almost never erred. However, on the same task in a group setting, where everyone else gave incorrect answers, nearly three quarters of people erroneously went along with the group at least once. People were defying evidence their own senses observed just for the reason of conformity. Choice architects can use this fact to move people in a better direction (Sunstein & Thaler, 2021). Social norms can influence higher or lower levels of waste sorting, energy consumption or mobility options. To reduce emissions, we must motivate people to shift to more energy-efficient cars and change modes of transport such as by using a bicycle instead of a car for short distances. One specific type of nudge here would be fitness challenges, where information about other people's cycling behavior is evaluated and then used by choice architects. A range of studies in the US, UK and Ireland have documented that social feedback combined with frequent information on energy usage of others, can reduce consumption by 7%. The use of social norms has been shown to be effective when peer comparisons are offered in combination with information on personal consumption behavior. The focus should be on situations where people have a personal point of reference (A. S. E. Nielsen, 2016). For water utilities, nudges can be cheaper and easier than building new dams, wells or plants. Startups such as WaterSmart, H2OScore and DropCount have developed tools which are using the human need for conformity to alter consumer behavior. For instance, they compare the water consumption of an individual with the usage of their neighbors (Wang, 2014).

2.7.4. Increase in ease and convenience

Resistance to change is often not based on disagreement or skepticism. It is often the perceived difficulty of a decision or the ambiguity of arguments which hinders people from making a good choice for themselves (Sunstein, 2019). When speaking of food, convenience is often associated with less healthy choices. Hanks et al. (2012) executed a study where convenience was associated with healthier choices. Healthier foods were made more convenient relative to less healthy foods. One of two lunch lines in a cafeteria was arranged this way and field researchers compared purchases and consumption before and after the conversion. The study provided evidence that the convenience line that offered only healthier food options nudged students to consume fewer unhealthy foods. Sales of healthy foods increased by 18% while the consumption of less healthy foods decreased by nearly 28%. Even small improvements of choice architects in making an option more convenient will have an impact. Experiments in Scandinavia have also shown that when meat-eating consumers are presented with menus that list vegan food at the top of the menu card, most will order vegan. The recycling tendencies of office workers suddenly rise if bins with visual signs are placed next to their desks, and the ease of videoconferencing tools has made us rethink flying (Tett, 2021).

> "My number-one mantra from Nudge is, Make it easy. When I say make it easy, what I mean is, if you want to get somebody to do something, make it easy. If you want to get people to eat healthier foods, then put healthier foods in the cafeteria, and make them easier to find, and make them taste better. So, in every meeting I say, Make it easy. It's kind of obvious, but it's also easy to miss" (Thaler, 2011).

2.7.5. Disclosure

Deliberately disclosing decision-relevant information in an explicit way can also be used to nudge humans towards better choices. Disclosures can be highly effective, but must be comprehensible and accessible to customers. One example of a disclosure nudge would be the communication of the environmental or economic impact of products or services (Sunstein, 2019). Another example of a disclosure nudge has been tested by Gimpel et al. (2020) to fight fake news on social media platforms. In an experiment the researchers simulated the Facebook newsfeed. In the nudging test, they disclosed related articles to the main article. The related ar-

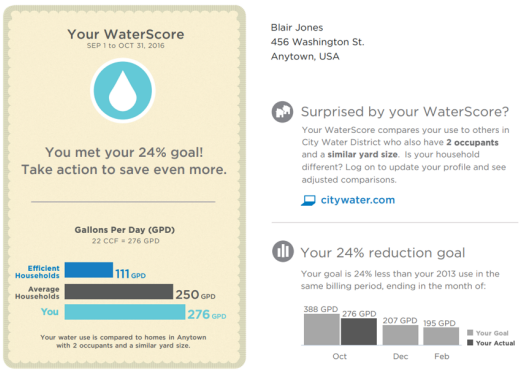


Figure 5: *Social norm nudges in a Generic WaterSmart Report* (Lade et al., 2020)

ticle provided a controversial view on the main topic. The study found that this disclosure nudge led to a positive influence on the detection of fake news on social media platforms.

According to Sunstein and Thaler (2021) an obvious nudge in the realm of climate change, would be a law requiring disclosure of all greenhouse gases by every country. Early progress on this initiative has already been made. The Paris Agreement requires a number of nations to disclose their emissions on a national level. A full disclosure via a greenhouse gas inventory would permit people to compare their emissions in detail and track changes over time. Seeing that list would lead to legislative initiatives and draw the attention of the media to the largest emitters by industry, region etc. While the inventory might not produce substantial changes on its own, it will nudge people towards better decisions by drawing attention to emission trends including which areas are of the problem.

2.7.6. Warnings - Graphic or Otherwise

Our attention is a scarce resource. Nudging can help us direct our attention towards a specific focus area (Thaler & Sunstein, 2021). When serious risks are involved, the best nudge might be a warning in large fonts, bold letters and bright colors to drive awareness. Warning nudges try to change the possible bad choice of a user by steering them toward an alternative and better option. A drawback of warnings is based on the human tendency towards unrealistic optimism. People might respond to serious threats by thinking that they will be just fine and discount the long-term disadvantages of their choices. In this case it makes sense to experiment with positive messages, providing some kind of reward for positive behavior which may not be monetary. Rewards in apps can nudge people towards better choices for themselves (Sunstein, 2019).

2.7.7. Precommitment strategies

People have diverse goals in their lives. They want to be more active, quit smoking, use less social media or engage in productive activities. To really achieve those positive changes, a useful nudge is precommitment strategies to engage in a certain action such as a work-out plan together with a fitness coach. People are even more likely to act in accordance with their goals if they commit to a specific action at a precise moment in time (Sunstein, 2019) With the 'Save More Tomorrow' plan, Thaler and Benartzi (2004) nudged employees to bear more responsibility for their own savings, by implementing a prescriptive savings plan. The employees had to commit in advance to allocate a portion of their future salary increases toward retirement savings. The intervention was effective as the savings rate of participants increased from 3.5% to 11.6% over the course of 28 months. Precommitment strategies can also help consumers nurture and facilitate their desire to a more environmentally and socially conscious lifestyle. To propel the sustainable transformation forward, consumers should commit to sustainable choices early on. Interventions could include asking them to sign a pledge or press a button to act in an environmentally-friendly way (DaSilva et al., 2022).

2.7.8. Reminders

Due to their limited attention, people sometimes simply forget something. Reminders are cues to pay attention and help us when we are absentminded. Technology has made it easier for choice architects to implement reminder nudges. With nearly everyone carrying a smartphone in their pockets these days, they can use it to send notifications in well-timed prompts. We get reminders of our restaurant reservations, our doctor's appointment or if a bill is due (Sunstein & Thaler, 2021). In a 693-person mega-study with Walmart, Milkman (2022) tested the effectiveness of text message reminders. The idea was to see which messages most effectively nudged people to get a seasonal flu vaccination. Patients who had previously opted in to receive texts from Walmart's pharmacy took part in the study. 22 different text reminders were tested, which increased vaccination rates by an average of 2%.

2.7.9. Eliciting implantation intentions

Implementation intentions can trigger numerous psychological processes which facilitate positive actions. These processes may relate to the anticipated situation or initiation of a goal-directed behavior (Gollwitzer, 1999). If someone elicits their implementation intentions, people are more likely

Figure 6: *Warning nudges on cigarette packages (U.S. Food and Drug Administration, 2019)*

Figure 7: *Example of a sustainability pledge in travel – Adapted from* Visit Iceland (2022)

to engage in activity. A simple question about future conduct can have significant consequences for decision behavior. Choice architects can use questions such as the below to direct people towards better decisions about health, wealth and happiness:

- Is protecting the environment important to you?
- Do you plan to protect your child with a vaccination?
- Do you plan to vote in the current election? (Sunstein, 2019)

2.7.10. Informing people of the nature and consequences of their own past choices

This tenth, important intervention in a choice architect's toolbox is similar to social norms, except the norm is directed to oneself (Wall et al., 2021). Private and public organizations have detailed information about the past behavior of people such as their monthly expenditures on their electronic bills. The problem is that individuals often lack this information. Giving them visibility of this information can nudge them to better choices (Sunstein, 2019).

Digital technology enables us to make this decision-relevant information available. By continuously monitoring users on activities and choice behavior we can nudge them to make better decisions for themselves (Karlsen & Andersen, 2019).

Thaler and Sunstein (2021) call this tool "smart disclosure". The data of credit cards, smartphones or other devices can be utilized for disclosure nudges, based on privacy-respecting, personal data. As in the WaterSmart example, it uses a social nudge as well as the smart disclosure of water usage. Another example would be using credit card usage to nudge users when something important is happening or the card is getting underfunded. A third example could be a child that has a nut allergy and you are out grocery shopping. The allergy could be contained in your shopper's club information. If you then buy the product which contained nuts, there would be a smart disclosure towards the customer.

Machine-readable information of past personal choices could even for the creation of 'choice engines' - technologies that interpret this data. Just as Netflix or Spotify help you decide which movie to watch or song to hear, choice engines can help you with decisions that have much higher stakes. They will enable consumers to take complex decisions such as finding the best mortgage or cellphone plan in a similar way as they search today for airplane tickets. Currently, it is not a lack of missing ingredients that has kept many choice engines from making the leap from beta testing to marketing disruption. It is easy access to data (Tucker & Thaler, 2013).

2.8. Ethical aspects of nudging

The high profile that Nudge received also led to criticism and ethical discussions on its techniques (Hallsworth & Kirkman, 2020). In 'The Ethics of Nudging', Schmidt and Engelen (2020) outline objections, including:

- Autonomy,
- Manipulation and dignity,
- Illicit ends,
- And structural reform instead of nudging.

First, some objectors worry the nudging undermines autonomy and volitional rational agency. However, nudges are used to closely align people's decisions to their ends and thereby strengthen their individual choices. Sunstein (2014) argues that many nudges are specifically designed to ensure informed personal choice. Preservation of freedom of choice is a vital part of the nudging concept as it guarantees autonomy. As in the GPS example we used to describe paternalism, nudges have the goal of increasing navigability. Nudges make it easier for people to get to their preferred destination, even when it is hard to navigate.

Second there is criticism that nudging is a manipulative tool. Some nudges obviously use manipulative techniques. However, if those interventions produce enough benefit, a minor amount of manipulation can be ethically justified. Painting dashed lines on roads manipulate people as they produce an illusion of speed, but they also reduce crashes. When choice architects use manipulative intentions which are against the overall benefit of a person, this would not be a nudge by definition and unethical behavior (Wilkinson, 2013). Choice architects must ensure that their designs are compatible with human dignity. If a person feels less healthy, sadder or poorer after being nudged, that is unethical behavior (Thaler & Sunstein, 2021).

The same is true for the third worry on the use of nudging. Many nudges are objectionable as the choice architect has the intention of perusing excessive paternalism and other illicit ends. Consider serious ethical issues if people use interventions to discriminate based on race, gender or religion. Even truthful information, for example, on crime rates by the government might fan flames of violence and prejudice (Sunstein, 2015). Nonetheless, in some cases the legitimacy of the goal is not granted. We all agree that people want to be healthier and not die in traffic, while other cases might raise stranger epistemological concerns (Schmidt, 2017).

The last critique on nudging is less directed on the technique itself, and more on its use in a real-world economic and political context. The objection is that nudging falls short on many big issues our society encounters. Obesity, climate change or consumer debt have a complex biological and social etiology. Nudging bears the risk of shifting the focus simply to individual decision-making without looking at the real underlying causes. However, there is no need to take a decision between nudging and other policies implemented by businesses or governments. They can tackle both those current challenges (Schmidt & Engelen, 2020).

The central arguments for the use of nudges are that both nudges and choice architecture are inevitable. Therefore, it is somewhat futile to discuss if their use is ethically right or wrong. Furthermore, these forms of choice architecture come in many forms and can even help us to reach ethical goals such as welfare, autonomy, dignity or other values. Any change in choice architecture must preserve freedom of choice, but still can run into serious ethical objections (Sunstein, 2015). Nudges are small design changes that can markedly affect individual decision behavior. This is why choice architects must ensure to not sway people to choices which they will later regret. They should have good reason to believe that the encouraged behaviors will improve the welfare of those being nudged. Nudges should never be misleading and people should preferably be able to opt out of the nudge with ease (Thaler, 2015).

2.9. Effectiveness of nudging

Nudges are almost everywhere in our lives today. If you are getting information on how many people booked a room in the last few hours, choosing a default option, or buying a product which has a label that provides proof of energy efficiency, you are being nudged. However, there are limits beyond ethical considerations on nudging, as in any tool for behavior change.

Oreopoulos and Petronijevic (2019) designed online and text message interventions for a sample of nearly 25,000 students to improve their college achievement. They found that these nudging tools had no significant influence on academic outcomes. Cantor et al. (2015) executed a study that examined how the display of calorie information on menu boards of fast-food restaurants in New York City changed consumer food choice. There was no statistically significant change over time in levels of calories or other nutrients purchased or in the frequency of visits to fast food restaurants. However, in research, knowing what does not work is as important as knowing which interventions are effective. Firstly, if choice architects are not able to nudge choosers towards a specific option, the findings could suggest that different and possibly more intensive interventions are needed. Secondly, even null results have implications as those nudges can be determined to not be used for a certain task or environment ("Nudges That Don't Nudge," 2020).

Overall meta-analysis on the effectiveness of the choice architecture method has proven to be effective in many cases. Szaszi et al. (2018) analyzed 2670 papers and found 93% of the studies contained at least one successful intervention. Hummel and Maedche (2019) calculated effect-sizes of nudges from different research areas. In the meta-analysis, nudges had a median relative effect size of 21% and an average relative effect size of 55%. They also split up the studies by context and nudge tool. Both tests showed that results vary based on context and nudging form where default options especially had larger median and average effect sizes than other categories. Precommitment strategies were found to be the least effective nudging tool.

In their meta-analysis, Hummel and Maedche (2019) also analyzed if there were differences between nudges in offline settings and nudges where IT was involved. They found the effect sizes of nudges in a digital setting were not different to the effects in conventional settings.

2.10. Digital Nudging

The digital ubiquity and information overload leads people to the limits of their cognitive processing capacity. This means in the digital space, we are often deciding based on system 1 – automated, fast and not consciously. The increasing use of digital technologies also means that many of our

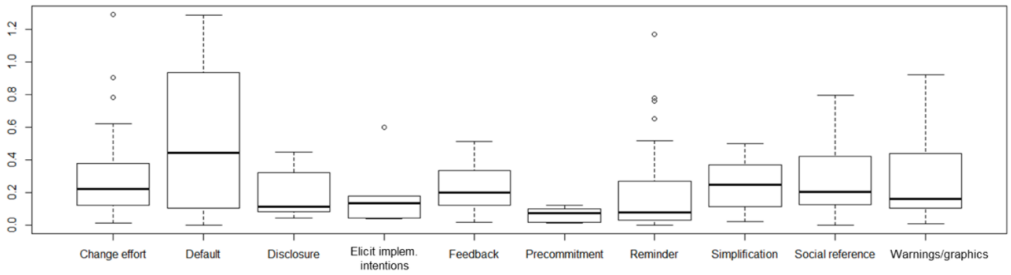


Figure 8: *Boxplot of relative effect sizes per nudging tool* (Hummel & Maedche, 2019)

choices are made in online environments today. User interfaces on our digital devices such as notebooks, smartphones or Internet of Things include choice environments ranging from e-government to e-commerce interactions. Even simple modifications of those environments can influence people's decision behavior. Digital nudging applies the nudging concept to user interface design elements in digital choice environments. It serves as a valuable tool for improving user experience and leading them towards more advantageous decisions for themselves without restricting freedom of choice. Compared to conventional forms of nudging, digital nudging also has the advantage of being easier, faster and cheaper to implement. In addition, user behavior can be tracked and analyzed with appropriate technologies to determine the effectiveness of nudges. One example for choice environments where digital nudges can be implemented are web-based forms (Mirsch et al., 2018; Schneider et al., 2018)

Digital nudging can change socially acceptable group behaviors and spread quickly throughout organizations. It also has the ability to test nudge campaigns in small segments and scale up what works best. A pilot can be a short, low-cost experiment (Dhar et al., 2017). Understanding digital nudges is important for the IT profession, as user interface designers create most of today's choice environments. User interface designers often focus primarily on usability and aesthetics, and neglect the potential of behavioral effects of different designs (Schneider et al., 2018). Nudging in a digital environment follows the same principles as in the physical world. The fundamental difference is the greater versatility and dynamic, which choice architects have in virtual spaces. It is a promising research field for scholars and practitioners in the field or in the areas of user interface, user experience and digital service design (Mirsch et al., 2018). Testing digital nudges is especially important as effectiveness is likely to depend on the context and goal of the digital nudge. While digital nudges work well in the context of hotel bookings, they might be ineffective in other domains. This may be due to different target users or the unique nature of the decision processes. Therefore, techniques such as A/B testing help choice architects find the most effective interventions (Schneider et al., 2018).

Web technologies allow us to track users in real-time and analyze their behavior. They also enable choice architects to generate information on the decision maker's personal characteristics and their environment. They have the ability to implement nudges by modifying a system's user interface. This can include actions such as setting defaults, displaying/hiding design elements or providing information on pledges. Notwithstanding the availability of dynamic adjustment on how options are presented, designers should follow commonly accepted design guidelines such as Apple's Human Interface Guidelines to ensure usability and consistency. If a particular Nudge does not produce the desired effect, choice architects should evaluate whether the Nudge is too obvious or not obvious enough (Schneider et al., 2018). Digital nudging can be used to tackle our current environmental issues. They provide ways to make pro environmental behavior less complicated for consumers (Zimmermann et al., 2021).

2.11. The importance of online experiments

People commonly assume the greater the investment, the larger the impact it will create. However, in the digital space, success is more about getting many small changes right to achieve progress. This 'learn fast' philosophy is based on the Lean Startup (Ries, 2011) where the premise is to get feedback early and validate that you are on the right path. It can be difficult to assess the potential of new ideas. Based on that, it is important to have the capability to run tests cheaply. The digital world is often viewed as turbulent and full of peril, but controlled experiments can help us navigate. They can show us answers, which were not obvious and provide proof on the value of an idea. Controlled online experiments can transform decision-making into a scientific, evidence-driven process rather than an intuitive reaction. They can help managers to make better decisions (Kohavi & Thomke, 2017).

> *"The premature outlay of huge amounts of money in pursuit of the wrong strategy is the thing to avoid. You need to have an experimental mindset." (Christensen, 2020)*

These tests can be run in a laboratory or in the field. Their purpose is to test a hypothesis in order to accept or reject it, based on measurable results. For any experiment there is no need for a fully functional app or website. A digital prototype acts as a proof of concept that allows testing before fully building the solution. Prototypes can be used to test the general feasibility or usability of a product and allow us to iterate, review, and refine big and small ideas. Some popular

Table 4: *Example applications of digital nudging – Adapted from* Weinmann et al. (2016)

Use case/ Information system	***Nudging example/behavior change intervention***
Business process management	Structuring complex input screens
E-business and e-commerce	Displaying limited room inventory during a hotel-booking process
E-finance and insurance	Setting defaults for frequently selected insurance plan options
E-government	Setting defaults to opt in for organ donation
E-health	Step counter app that provides feedback on activity levels
E-learning	Reminder to learners to engage with course content
Green IS	Smart meters to encourage energy savings
Security and privacy	Displaying the strength of selected passwords
Social media	Giving incentives, such as badges, for sharing or other activities

tools for rapid prototyping include Adobe XD, Figma, and Sketch (Levy, 2015).

2.12. A/B Testing

Once thought to be exclusive to large technology firms, A/B testing has become a viable and cost-effective way to identify and test value-creating ideas for many businesses. In statistical terms, it is a method of two-sample hypothesis testing. The A and B method are compared by outcomes of a controlled experiment. What sounds like a simple test of two variants has been a revolutionary approach to decision-making driven by data. This iterative approach is a key success factor for many Silicon Valley giants that have a deeply rooted so-called 'testing culture'. A/B tests enable companies to advance decision-making from gut feeling, assumptions or a battle of long-winded arguments to facts based on statistics. What has not been tested and thus proved or disproved is only a personal opinion. The method is often discussed in the context of digital marketing, web design or other marketing or technology-focused applications. A/B testing methods are sometimes referred to as split testing (Vallee, 2016; Witzenleiter, 2021). Fung (2017) who founded the applied analytics program at Columbia University, explained A/B testing as a way to compare two versions of something to figure out which performs better. A simple example of an A/B test is the size of a subscribe button on a website. In this case, the metric to evaluate performance is the number of visitors who click on the button. To run the A/B test, two groups of randomly assigned users visit two different versions of a website. The only difference between both sites is the size of the button, which enables an analysis of which button size caused more visitors to click. However, sometimes other variables such as mobile versus desktop, will drive results on average. While randomization is an important part of testing, this may result in set A containing slightly more mobile users than set B, which may cause set A to have a lower click rate regardless of the button size users are seeing. Statistical blocking should be used in this example by first dividing users by mobile and desktop and then randomly assigning them to each version.

On websites, A/B tests can be run on diverse elements such as headlines, call to action text, call to action location, pop-ups, featured images or copy. Each and every design element which can be changed can also be tested. Accurate testing can drive a substantial effect on the effectiveness to achieve a certain goal. There is a possibility that one variation can work two, three or even four times better, with only minor investment or effort (Pateil, 2022). The power of A/B testing is that it provides evidence to understand why something happened. In other words, it lets us establish causality by making a change and measuring its effect compared to a control group. In addition to statistical significance, effect sizes can be calculated between the control and experimental group. Understanding these causes of behavioral effects can help designers explore what will happen if they make changes in a user interface. This enables us to understand with accuracy how design changes cause changes in our users behavior. Furthermore, A/B testing can protect companies against the very human tendency to see patterns in data and behaviors that confirm what we already think, and mitigate the risks of investing time and resources on assumptions that are not proven (King et al., 2017).

Before comparing the A version and the B version against each other, it is important to decide which score will be measured as a success factor. To start an A/B test effectively, there must be a clear understanding of the overall purpose of the website or one quantifiable, specific element which should be improved after the testing process. E-commerce businesses could define revenue per visitor as a success metric, fundraising sites could use completed donation forms. or Google's search engineering team could use abandonment rate as a metric, which indicates when a user leaves a search results page without clicking anything. Without a clear quantifiable success factor, it is tempting to focus on vanity metrics, which are not relevant to the overall goal (Siroker & Koomen, 2015).

3. Methodology

The function of all science is to investigate answers to questions with as much objectivity, ethical diligence, and rigour as possible (Jackson et al., 2007). Certain types of research questions call for specific approaches. Researchers should use a qualitative approach if the defined problem refers to:

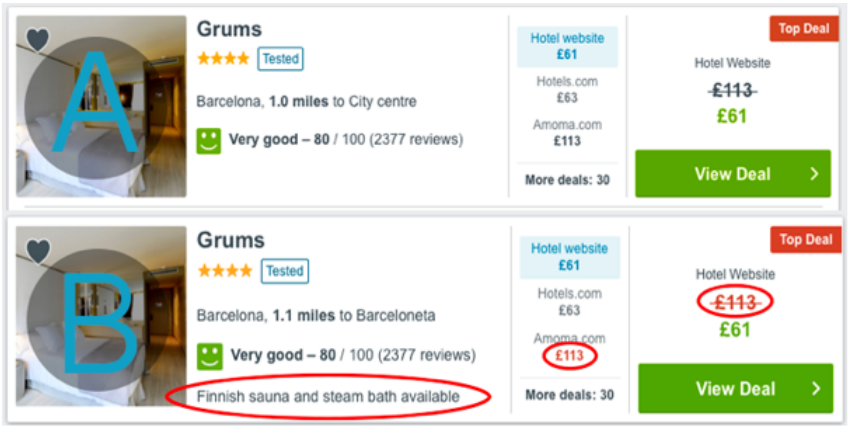


Figure 9: A/B Testing on the metasearch platform trivago (Satouri, 2019)

Table 5: *A/B Testing goals per site type – Adapted from* Siroker and Koomen (2015)

Website Types	Common Conversions & Aggregate Goals
E-Commerce Selling products online	– Complete purchase – Products added to cart – Product page views
Media/Content Focused on article or other content consumption	– Page Views – Articles read – Bounce Rate
Lead Generation Acquiring new business through name capture	– Form Completion – Clicks to a form page (links may read “Contact Us”)
Donation Sites aiming to collect donations	– Form completion – Clicks to a form page (links may read “Donate Now”)

- The identification of factors that influence an outcome
- The utility of a planned intervention
- Or understanding the best predictors of outcome.

On the other hand, if little research has been done on a certain topic and the merits of a concept or phenomenon need to be explored, a quantitative approach should be used. These approaches are plans and procedures of data collection, analysis and interpretation. Quantitative research statistically assesses some aspect of a research problem through the use of experimental or survey designs. Experimental research seeks to determine if a specific treatment of one group creates different outcomes from another. In those experiments, hypotheses are often used to answer research questions. They help to focus on the purpose of the study and inquire about the relationship of variables (Creswell, 2009).

Chambliss and Schutt (2016) also emphasize that experimental research provides the most powerful design for testing whether an association exists between an independent and dependent variable. These causations are then tested. One group receives some ‘treatment’ which is a manipulation of value of the independent variable. The other group is termed the control group and does not receive that treatment. As a simple example, consider a research question of whether drinking coffee improves one’s writing of a Master’s thesis. To test this question in an experiment, two groups would be compared. One group would be comprised of participants who always drink two cups of strong coffee while writing the thesis and the second group would be the control group comprised of people who will not drink coffee. At the end, all cases will be compared based on which group received the better grading. To test these or any other causal relationships, experimental research has three common features:

1. ***Two comparison groups*** (Experimental and control group)
2. ***Variation in the independent variable*** before assessment of change in the dependent variable, to establish time order
3. ***Random assignment*** to the two (or more) comparison groups

This thesis aims to understand the utility and effectiveness of digital nudges in a quantitative, experimental research approach. Specifically, this study aims to examine if nudges can effectively influence decision behavior towards

more sustainable choices in a travel context. Based on findings from current literature, we already know which travel activities cause greenhouse gas emissions, how and why nudges are working, as well as which nudging tools are available to choice architects. Additionally, current literature indicates the potential to nudge people towards more sustainable options in a travel context and on digital channels. Therefore, it is expected that the implementation of digital nudges has a causal relationship to the choice of the most sustainable travel products.

3.1. Hypothesis Development

The term hypothesis has already been mentioned in this thesis. In general, hypotheses are predictions that researchers make about the expected outcomes of a relationship among variables. Hypotheses make specific testable links between theories and their measurement. The goal is to form this information into a predictive statement. These statements are tested and may be confirmed, partially confirmed or proved false (Creswell, 2009; Williams et al., 2021). Going back to the coffee drinking experiment example, the test hypothesis for this research question would be: Coffee drinking improves the writing skills of students writing their Master's thesis.

These hypotheses reflect the purpose of the study. In the absence of any evidence to the contrary, the simplest starting point for researchers is to assume there is no relationship between the dependent and independent variable. This defines the null hypothesis. It is important to note that testing hypotheses has nothing to do with what the researcher wants to be true. It simply reflects an agnostic position based on the data of two samples. In order to test any quantitative hypothesis, measurable variables are necessary. The data must be generated by the same process before comparing them in statistical tests (Easterby-Smith et al., 2015).

In this experiment, the null hypothesis (H0) represents there is no relationship or significant difference between the group which received specific Nudges and the control group which received no Nudges. However, the theoretical background posits a causal effect of Digital nudges on the choice of offers, which emit the least greenhouse gases such as CO2. This causal relationship is formulated in the alternative hypothesis (H1). The literature meta-analysis by Szaszi et al. (2018), Mertens et al. (2022) as well as Hummel and Maedche (2019) provides evidence that Nudges are effective in certain contexts such as health, energy and finance.

> *H1: Digital nudges lead to the booking of the most sustainable travel offer*
>
> *H0: Digital nudges do not lead to the booking of the most sustainable travel offer.*

3.2. Study Design

To obtain the data and test the hypothesis afterwards, this experiment uses two sample groups which are compared against each other. In an online travel booking process, one group receives the regular process without any nudging intervention (G1). The other group (G2) receives one specific Digital nudge in the same booking process. To test the effectiveness of different forms of Nudges, various interventions are tested in G2 and compared against the control group, G1.

According to Mirsch et al. (2018) Digital nudges must always be developed and tested for a specific application context and should not be considered as best practices without reflection. They developed a systematic approach to design effective and user-centric Nudges at the Competence Center Digital Service Innovation at the University of St. Gallen. Going beyond a solely trial-and-error procedure when designing user interfaces, it aims to avoid unnecessarily long test and evaluation cycles. The model is used in this experiment to test the hypothesis and evaluate the effectiveness of different nudging tools.

3.2.1. Definition and Analysis of the Digital Nudging Environment

In the first phase of the systematic approach by Mirsch et al. (2018) specific goals of the interventions should be defined. In the case of this research project, the goal is to move consumers to choose sustainable options when booking their holidays. Furthermore, it is necessary to select the examined user interface in the first step. This is important as mobile applications for example have different design guidelines, strengths, weaknesses and requirements for designing digital nudges. This research is limited to mobile testing as according to Arora (2021) it is the primary device for travel planning in the majority of the world and mobile booking is set to soon surpass desktop booking in volume.

After defining the goal of the intervention and the user interface, the desired behavior is determined. It states which decision behavior is expected based on the Nudge. The nudged and desired behavior in this thesis is always the option that emits the least greenhouse gases. As a result, there are two possible outcomes per decision:

- The user chooses the option with the least greenhouse gas emissions. In the case of G2, this option includes one specific Nudge.
- Or (0) the user chooses an alternative option. This is any option which is not the most sustainable travel offer and is not designed with a Digital nudge.

Before implementing Nudges and testing their effectiveness in a travel context, it is necessary to develop the travel booking process. The user interface design process starts with developing the general version without Nudges. This is the version which is later tested by the control group G1. The experiment uses a simulated travel booking process where users can choose different services. These offered services have different levels of greenhouse gas emissions.

Based on the emissions analysis of the travel industry in the theory section, experiment participants receive 4 differ-

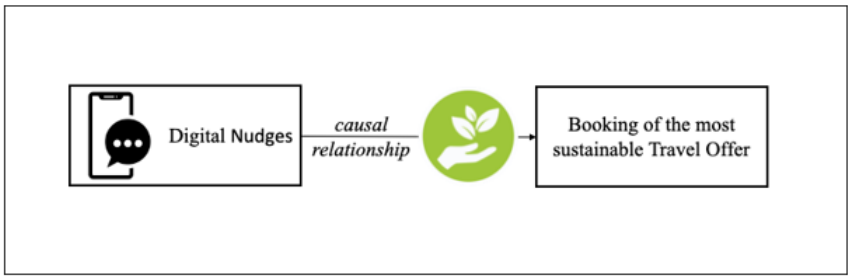


Figure 10: *Hypothesis of Decarbonizing Travel Decisions by Using Digital nudges*

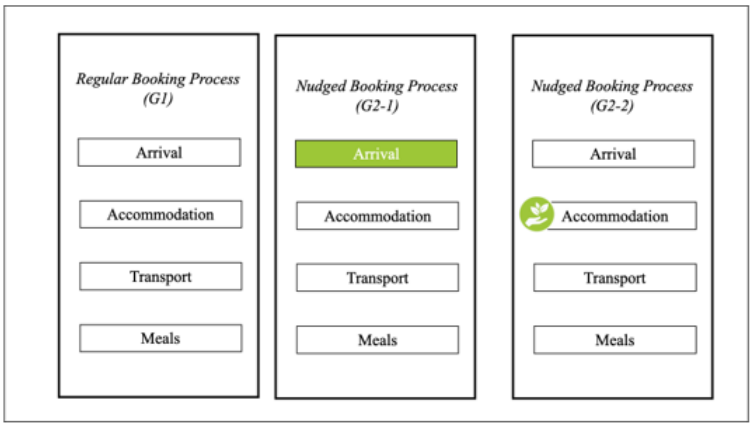


Figure 11: *Regular and nudged booking process*

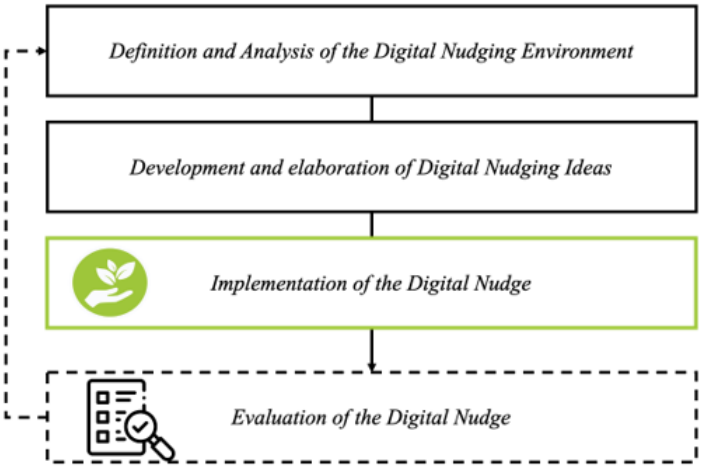


Figure 12: *Systemic approach for Digital Nudging – Adapted from* Mirsch et al. (2018)

ent booking questions. Users can choose freely for their preferred option in the realm of transport, accommodation and restaurants.

- ***Arrival*** – How to get there?
- ***Accommodation*** – Where to stay?
- ***Transport*** – How to get around?
- ***Meals*** – What to eat?

Transport causes the majority of greenhouse gas emissions and thereby is the most powerful lever in sustainable behavior that is included with two decisions. Arrival asks participants to choose their preferred option to travel to the desired destination from their starting point. The transport decision is focused on local mobility options at the destination.

Additionally, the booking process also includes two questions in the hospitality field. Users can select one of the offered accommodation options. The last question relates to

which meals guests want to have during their stay. This question is formulated in a generic way as the decision can relate to the booked hotel, but also other aspects of the journey such as individual restaurant visits. Shopping, which is another significant cause of CO2 emissions in the travel industry, is not covered in the booking process. This is due to the fact that shopping decisions take place during the journey and are not pre-booked on digital channels.

The environment of our experiment is a simulated booking process. It can be defined as a laboratory experiment in the digital space to test the hypothesis. Although researchers conduct experiments in various settings, laboratory or simulated domains have unique advantages including the ability to create and simulate artificial conditions, direct comparisons, replications, and measurement technologies. They allow researchers to build the necessary conditions for hypothesis testing and provide causal inference. Furthermore, laboratory experiments can simplify complex theories about human behavior, communication, and perception. Beyond those advantages of the simulated booking process, critiques of laboratory experiments stress the associated disadvantage of lacking external validity. The sterility of the approach is criticized too as it provides situations which are too abstract and differ too much from the real-world decisions that individuals make. Therefore, researchers should use field context rather than abstract terminology in their experiments. In search of greater relevance, this enables preventing unnatural behavior of experiment participants in a controlled environment (Allen, 2017; Harrison & List, 2004).

To provide users with context in this research project, the experiment starts by setting the scene. Users are asked to book a journey from Munich to Milano. Their task is to decide on their desired accommodation, transport and food options during their journey. The trip will start in Munich at noon on Friday, 12th August. The introductory remarks also include the information that they will travel with one companion and will return to Munich on the evening of Sunday, 14th August. The participants of the experiment are asked to choose the options that suit them best. The data on prices, routes and times was also calculated for the same dates and accessed on 16th July 2022 to make the simulated booking realistic.

The first question arrival refers to the transport from Munich to Milano. For the simulated booking process, we need the duration of the option and the prices. Additionally, the total emissions were calculated to gain information on which option is the most sustainable. Additionally, the total emissions are needed at a later time for the disclosure nudge. Information on prices, distance and duration were ascertained using Google Maps, Google Flights, Flixbus and Deutsche Bahn. The total average emissions are according to the Umweltbundesamt (2021) and include CO2, CH4 and N2O. These emissions are declared in a CO2 equivalent.

The price of travelling to Milano per car is not calculated as the related costs depend on the car ownership of the user. The journey by bus is the option that creates the least emissions – 13.635 g CO2 equivalent. As the most sustainable mode of transport, it is the option which will later be nudged. All emissions were calculated for the outbound and return journey. In addition, the indicated prices are calculated for the entire trip. The cheapest economy fare has been chosen for all modes of transport, including special discounts such as early booking deals.

After choosing the mode of transport. the experiment participants are asked to select their accommodation. Data of a Booking.com search from 16 July, 2022 is used with 4 different hotel options which users can choose from. The indicated price per person and night was calculated by using the lowest available rate for the cheapest room of the hotel on the booking platform. There is no data on which hotel has the lowest greenhouse gas emissions per person. The assumption to test the effectiveness of nudging one specific hotel is that Westin Palace is the most sustainable accommodation with the lowest CO2 output per person.

Another transport decision relates to the selection of mobility options at the destination. Users can choose between (1) public transport, (2) taxi / ride hailing or (3) taking a rental car. As this question is a general one, there are no price indications for the different options. If participants want to take the most sustainable choice, they would select the public transport option.

The last step of the booking process is the selection of food options that users plan to consume during their stay in Milano. Scarborough et al. (2014) estimated the difference in greenhouse gases of different dietary options. Users in the experiment can decide between a (1) vegan, (2) vegetarian, (3) high-meat based and (4) medium meat-based diet. The vegan diet is the most environmentally sustainable one. Guests who choose this option have mean greenhouse gas emissions of 2,890 grams of CO2 equivalents per day based on a 2,000 kcal diet.

Following the definition of decision areas and choice options, the regular booking process can be designed. The user interface design for the experiment was created with the web-based design tool Figma. It is a free, intuitive and user-friendly tool to create designs.

3.2.2. Development and elaboration of Digital Nudging Ideas

In the second phase, Mirsch et al. (2018) describe the importance of dealing intensively with the effects and implementation opportunities of Digital Nudging. The aim of the approach is to gain an in-depth understanding of which nudges could be used to achieve the goal of the interventions. Subsequent prioritization makes sense in order to further specify the most promising Nudges. The specification can be made by prototyping the different nudging ideas.

As described in the theoretical background, there are various ways to categorize different nudging tools. This thesis uses the framework by Sunstein (2019). The same categories were also used in the literature meta-analysis by Hummel and Maedche (2019).

The overall goal is to test every nudging tool out of our toolbox. This enables us to identify the most promising nudging ideas, which will be tested for effectiveness. For this pur-

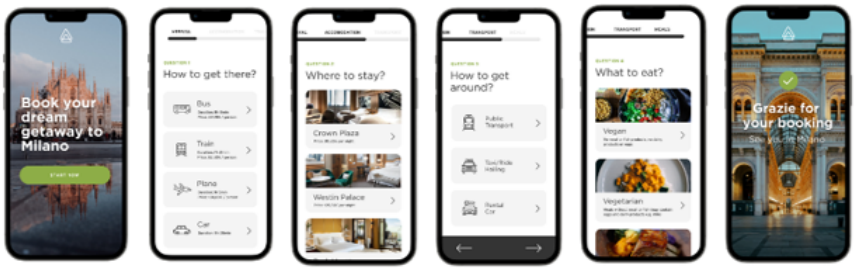


Figure 13: *The regular booking process (G1)*

Table 6: *Modes of transport – Sources (Deutsche Bahn, 2022; Flixbus, 2022; Google Flights, 2022; Google Maps, 2022; Umweltbundesamt, 2021)*

Mode of transport	*Distance*	*Average Emission*	*Total Emission*	*Duration*	*Prices*
Bus	505 km	27 g/Pkm	**13.635 g**	8h 5min	69.98 EUR
Train	388 km	50 g/Pkm	19.400 g	7h 21min	82.80 EUR
Car	495 km	152 g/Pkm	75.249 g	6h 20min	
Plane	348 km	284 g/Pkm	98.832 g	1h 5min	180.00 EUR

Table 7: *Accommodation Options - Source* (Booking.com, n.d.)

Accommodation / Hotel		*Price per person / night*
Crown Plaza	https://www.booking.com/hotel/it/milan-city.de.html	80.25 EUR
Westin Palace	https://www.booking.com/hotel/it/westinpalacemilano.de.html	139.50 EUR
Park Hyatt	https://www.booking.com/hotel/it/park-hyatt-milano.de.html	423.00 EUR
Mandarin Oriental	https://www.booking.com/hotel/it/mandarin-oriental-milan.de.html	747.50 EUR

Table 8: *Dietary Options and emissions - Source (Scarborough et al., 2014)*

Dietary Option	*Mean Emissions per 2,000 kcal diet*
(1) Vegan	2,890 g
(2) Vegetarian	3,810 g
(3) Meat Lover	5,630 g
(4) Standard Diet	7,190 g

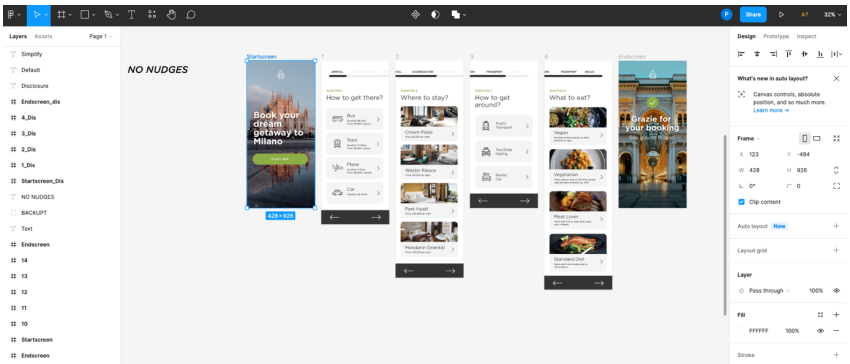


Figure 14: *Designing the booking process in Figma (Screenshot)*

Table 9: *Overview of different (Digital) Nudges – Adapted from* Sunstein (2019)

(Digital) Nudges	*Intervention implementation*
Default rules	Preselecting a choice option
Simplification	Reducing complexity of a topic
Use of social norms	Providing information about the decisions that others made
Increase in ease and convenience	Reducing barriers of a choice
Disclosure	Providing relevant, comprehensive information
Warnings, graphic or otherwise	Alert people about serious risks
Precommitment strategies	Commitment to a certain course of action
Reminders	Sending reminders (for example, by e-mail)
Elicitation implementation intentions	Asking persons for intended actions
Informing people of the nature and consequences of their own past choices	Disclosing previous personal choices

pose, the general booking process version without Nudges will be modified. In Figma the general version is duplicated and adapted with one specific Digital nudge. Based on the hypothesis, this Digital nudge will cause more users to choose the most sustainable option.

3.2.3. Default nudge (G2-1)

The first Digital nudge implemented is a default nudge. It is a very simple adaptation of the general booking process. When users choose their preferred mode of transport from Munich to Milano, the nudged option would already be preselected. If users do not proactively opt for another choice, they would be choosing to go by bus by default, which causes the least greenhouse gas emissions.

3.2.4. Simplification nudge (G2-2)

Even if users want to choose a sustainable option, it is difficult to find out which decision will have the lowest emissions. Furthermore, most of us cannot relate to a certain specification of CO2 equivalents. With the simplification nudge, users can easily see which decision can help to protect the environment. In the digital experiment, the green hotel label flags accommodations that are taking significant steps to reduce greenhouse-gas emissions and make the guests' stay more sustainable. To test the Nudge, the assumption made is that the Westin Palace Hotel is part of this program and the most sustainable option for the participant.

3.2.5. Increase in ease and convenience nudge (G2-3)

For the next nudging tool, the user interface is redesigned in a way that makes it easier for users to choose the most sustainable option. When asking users to book their local transport option in Milano, the regular booking process has the 3 options presented in the same way. To increase ease and convenience of choosing the most sustainable option, public transport is highlighted and complemented with a picture, while the alternative two options are less conspicuous.

Furthermore, the increase in ease and convenience nudge is not limited to the design of user interfaces in a simulated booking process. These Nudges can be about more than just making the most sustainable option more attractive visually. What (Thaler, 2011) described with "Make it easy" can be achieved by making the actual travel service more convenient and better for consumers. This means improving prices, quality and convenience of the most sustainable travel offer. For example, the train connections from Munich to Milano could be made less costly, faster and more punctual which makes the choice easier and more convenient for guests.

3.2.6. Social norm nudge (G2-4)

It is expected that most users will choose their usual diet in the regular booking process. Based on the influence of others, a social norm intervention could nudge more users to choose a vegan diet during their holidays. The design of the nudged version would indicates that 64% of other guests choose to eat vegan dishes during their stay. A it's popular sign design intervention is implemented at the vegan diet option with additional information on the bottom of the interface.

3.2.7. Disclosure nudge (G2-5)

Another nudge to choose the most sustainable option to travel from Munich to Milano discloses the weight of CO2 equivalent greenhouse gas emissions for the different transport options. It transparently and objectively communicates the environmental impacts of the different options. As an additional visual element, the different CO2 outputs are indicated with green, grey and red. The green option of travelling by bus is the nudged element with 13.635 g of emissions.

3.2.8. Warning nudge (G2-6)

The disclosure of CO2 equivalent emissions is used for the next nudge. As stated in the theoretical background, warnings can be addressed in a positive or negative way. The nudge used in this experiment is a positive example of mobilizing people towards a common goal of fighting the current climate crisis and addressing the individual power of the chooser. The G2-5 version also uses the developed G2-4 nudge, as users may not be familiar with which option actually is the most sustainable one. As in the previous option,

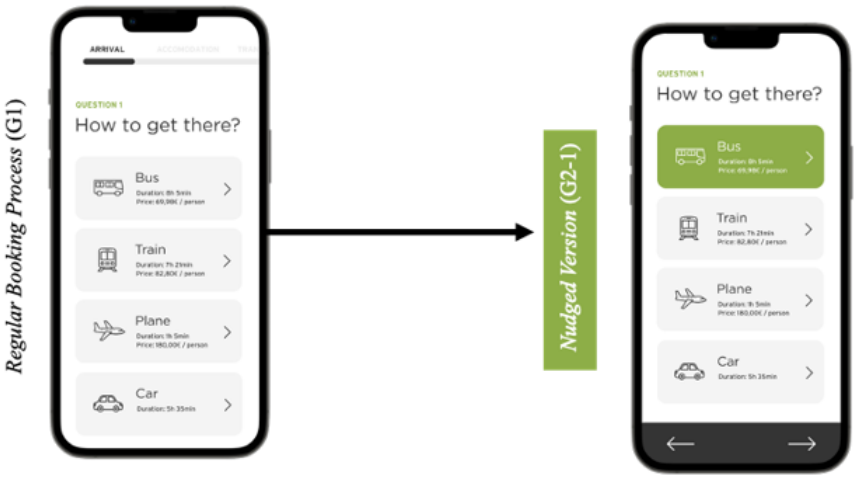


Figure 15: *Default rule nudge G2-1*

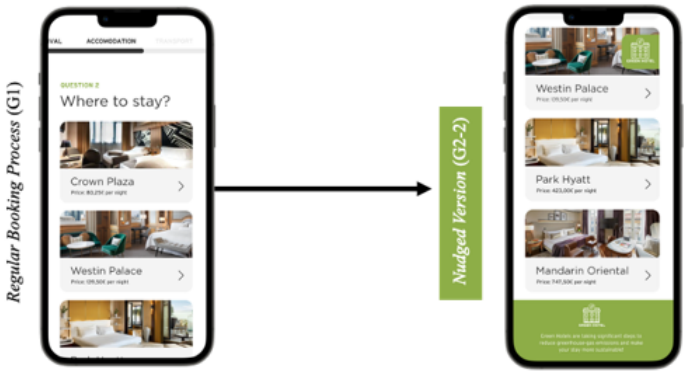


Figure 16: *Simplification nudge G2-2*

Figure 17: *Social norm nudge*

the effectiveness is measured by how many people choose the most sustainable option in comparison with the regular booking process.

3.2.9. Precommitment nudge (G2-7)

The precommitment nudge is implemented before the start of the actual booking process. Choosers can commit or not commit to being a sustainable traveler. A commitment is a pledge to act responsibly by choosing offers that emit less greenhouse gases into the atmosphere. Similar to

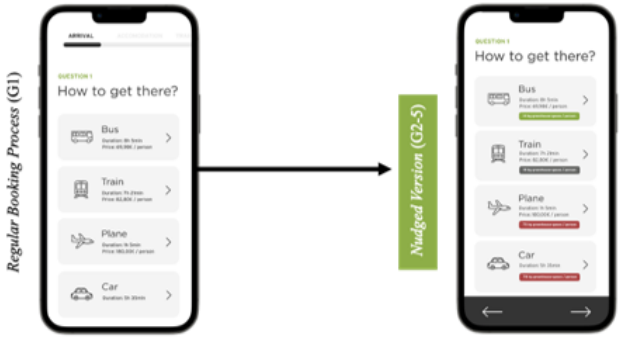


Figure 18: *Disclosure Nudge*

Figure 19: *Warning nudge*

the warning nudge, another nudged version is used to guide users about which option is the most sustainable choice. Both cases can be defined as a hybrid nudging technique.

3.2.10. Reminder nudge (G2-8)

As our regular process design is simulated, it is not feasible to set reminders automatically after the completion of the booking. However, to explore the capabilities of reminders to nudge users towards more sustainable products, an e-mail reminder is used. As the journey is planned to start on Friday, 12th August 2022, users will receive an e-mail which provides them with the option to change their decision three days prior to their departure. Users who did not choose the climate-friendly vegan food option will be nudged to click a link.

3.2.11. Informing people of the nature and consequences of their own past choices

This nudging tool uses the past behavior of people as a baseline. This personal information is not available in our experiment and therefore the nudge cannot be tested for effectiveness in a travel context. For future research in this realm, an existing booking platform could use the past behavior of users obtained from customer relationship management tools. One practical instance of this Nudge would be by indicating if the person's CO2 output is higher than the average user.

3.2.12. Elicitation implementation intentions nudge (G2-9)

The last Nudge tested in the experiment elicits user intentions to choose the most sustainable option. Experiment participants are asked if protecting the environment is important to them before the actual start of the booking process. Similar to the warning nudge, this idea builds up from the G2-4 nudge. This is important as users have to be aware of which is the most sustainable option after they agree to act sustainably. The effectiveness of this hybrid nudge will be measured by how many people choose the low-emission bus option.

3.3. Implementation of the Digital nudge

The third phase of the systemic digital nudging approach by Mirsch et al. (2018) is implementation. Following the careful definition of the goals, understanding the users and their environment, and the development of nudging ideas, this is the last step before testing.

In the implementation phase, Digital nudges are transferred to the corresponding decision-making environment or the user interface. For this research project, the Figma designs are directly used for user interface testing. These designs are imported to an A/B testing tool, which is the decision-making environment of all participants. Maze is used to conduct the A/B testing as it has a seamless functionality to import Figma links. Every nudge is set up as

Figure 20: *Precommitment nudge*

Figure 21: *Elicitation implementation intentions nudge*

one experiment in Maze. Wiggers (2022) describes Maze as a product research platform that facilitates tests and surveys. The tool enables digital marketers to observe how users interact with a product and generate reports. Maze can generate sharable links with instructions which are used to enroll participants to the testing process. It also includes a feature that enables researchers to capture video and screen recordings of people testing the products. In addition, Maze also offers mobile testing on desktop devices or tablets. This provides the experiment with the opportunity to test on all devices for a smartphone user interface.

The following overview shows the experiments created in Maze. The nudged decision is the part of the booking process with an intervention. The results will be analyzed whether a user chose the most sustainable option (1) or chose an alternative (0). The nudged versions will later be compared against the regular booking process.

In addition to the A/B testing, Maze is also used to obtain personal information of users to analyze the sample of the experiment. Experiment participants are asked for their:

- ***Age –*** *What is your age?*
- ***Gender –*** *What is your gender?*
- ***Country of Residence –*** *What is your country of residence?*
- ***Importance of Climate Change –*** *How important is the issue of climate change to you?*
- ***and E-Mail Address*** *(for the reminder nudge) – What is your e-mail address?*

To test the research design, a pretest was carried out with 20 participants. The pretest asked participants for qualitative feedback on the process and also used the screen recording function of Maze to identify barriers in the simulated booking processes. Minor adjustments to the experiment were made before the actual sample and data collection.

3.4. Sample & Data Collection

According to a publication by Allen (2017) true experimental designs are characterized by the random assignment of participants to experimental conditions. This provides researchers with the advantage that causal relationships can be clearly demonstrated. Creswell (2009) also specifies that if one of the groups receive a treatment and the other group does not, researchers can observe whether it is the treatment and not other factors that influence the outcome. Ideally, each individual in the general population has an equal probability of being selected for the experiment. On its Experimental Design Website, Yale (2022) mentions randomization in experiments as a common practice for researchers as it is

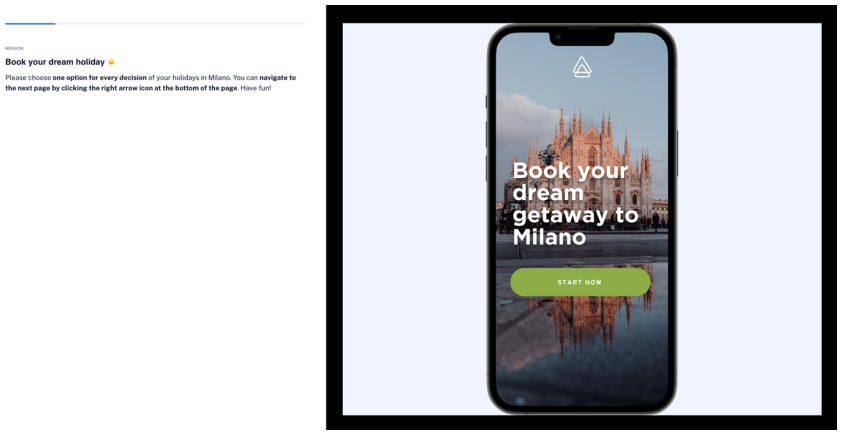


Figure 22: *The Maze experiment conducted on a desktop PC (Screenshot)*

Table 10: *Overview of Maze Experiments / Nudges*

Experiment / Nudge		*Nudged Decision*	*Nudged Item*
G1	Regular Booking Process	-	-
G2-1	Default nudge	Arrival	Bus
G2-2	Simplification nudge	Accommodation	Westin Palace
G2-3	Increase in ease and convenience nudge	Transport	Public Transport
G2-4	Social norm nudge	Meals	Vegan
G2-5	Disclosure nudge	Arrival	Bus
G2-6	Warning nudge	Arrival	Bus
G2-7	Precommitment nudge	Accommodation	Westin Palace
G2-8	Reminder nudge	Meals	Vegan
G2-9	Elicitation implementation intentions nudge	Arrival	Bus

the most reliable method of creating homogeneous treatment groups without involving any potential biases or judgments.

Overall, this research project tests 9 different nudges for effectiveness. 8 nudges are directly applied in the booking process. The reminder nudge is the only intervention that is applied afterwards on a different channel. Additionally, there is the regular booking process which is used to test the hypothesis. This means our experiment has 9 different Maze sharing links.

To randomize our sample test, participants will receive one link which evenly distributes the traffic to the 9 Maze experiments. This process is done automatically by using the online tool Linkly. In addition to splitting, the links can also be analyzed constantly while the experiment is running. This enables traffic monitoring on the link and the analysis of which sources are the most promising ones, which users are clicking the link, etc.

The nudged versions will receive 11% of the traffic and the regular booking process 12% of the volume. By virtue of this procedure, the sample is randomized and as traffic is split equally, it can be inferred that the number of participants is also evenly distributed. Upon completion of the experiment, the reminder nudge will be tested by e-mail. Users that have not chosen the most sustainable option in G1 will receive this specific intervention.

As part of an online experiment, Budiu and Moran (2021) from the renowned user interface and user experience consulting firm Nielsen Norman Group, defined 40 participants as an appropriate number for most quantitative studies. This sample size will typically lead to a trustworthy prediction for the behavior of your overall population if researchers try to measure binary metrics such as success rates or conversion rates. Experiments with 40 or more participants will produce results with a small margin of error and a high confidence level.

The goal of this research project is to attract 50 participants for every experiment. This should reduce the risks of the findings not representing the behavior of the user population. For the reminder nudge, users who did not choose the option with the least CO2 emissions, will take part in two experiments. A 300 EUR Amazon voucher was used to increase participation rates for the online experiment by incentivizing users to take part in the experiment.

4. Analysis

Overall, 456 participants completed the entire booking process of the nudging experiment. The test persons were recruited via social media networks with Instagram being the most important one (38%). The link to the experiment's random link rotator was shared along with social media creatives with personal contacts and travel influencers. The strategy

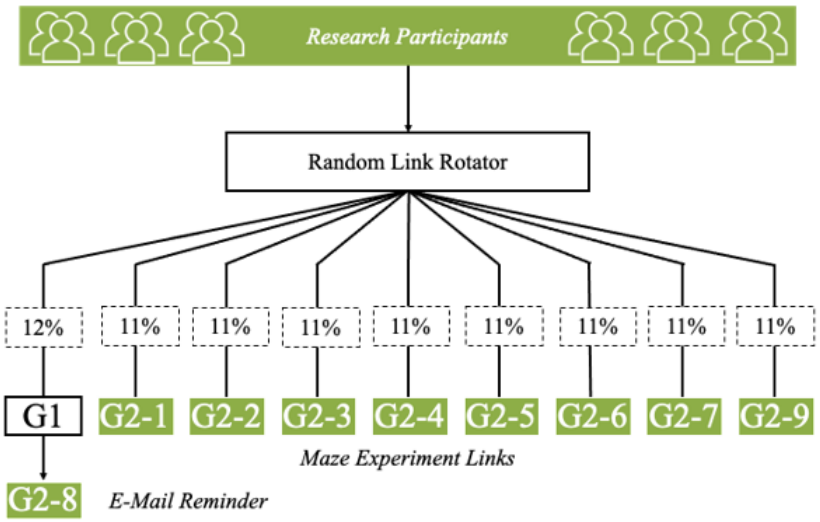


Figure 23: *Randomization in the experiment*

to activate diverse networks at the same time proved to be very successful. After major social media influencer accounts such as @manueldietrich or @evolumia posted the experiment, a major increase in experiment participants was observed. Furthermore, personal social media accounts were used to call attention to simulated booking process. The social media postings attracted 356 participants for the research project. To complete the sample, 100 more testers were acquired from the integrated panel function in Maze. The experiment started with the first testers on 29th July 2022 and ended on 11th August 2022. The dataset was exported from Maze and cleaned in Excel. The following analysis has been performed using SPSS.

The median age of experiment participants was 28 years with the sample reaching an average of 30.17 years. The sample consisted of participants identifying themselves as 226 female, 219 as male and 1 as non-binary. Looking at the countries of residence, the online booking process was completed by users from 29 nations, with the 5 main ones being:

- **Austria** – 175 participants ***(38.38%)***
- **Germany** – 124 participants ***(27.19%)***
- **United Kingdom** – 31 participants ***(6.80%)***
- **United States** – 30 participants ***(6.58%)***
- **Italy** – 30 participants ***(6.58%)***.

Participants were also asked to state how important the issue of climate change was for them on a scale from 1-10. 1 was defined as not important and 10 as very important. As demonstrated from the results, the sample was very aware of global climate challenges with 76% of surveyed participants rating the issue between 8 to 10.

Testers mainly used iOS devices (207) to perform the simulated booking process. Other operating systems which we identified in the sample by using Maze were Android (106), Windows (104), Mac OS (37), Chromium OS (1) and Linux (1).

4.1. Binary Logistic Regression

To analyze the effectiveness of Nudges, this research project uses Binary Logistic Regression. The goal for any regression model is to find the best fitting, simplest model to understand the relationship between the Ys and the Xs, and to be able to determine appropriate statistical conclusions from data (Fritz & Berger, 2015). According to the method consulting at the University of Zürich (2022) Binary Logistic Regression examines the relationship between the probability of a dependent binary variable taking the value of 1 and one or more independent variables. This means that it is not the value of the dependent variable that is predicted, but rather the probability that the dependent variable will have the value 1. To use this regression model, the dataset must fulfill the following prerequisites:

- The dependent variable is binary (0-1);
- The independent variables are coded scale or, in the case of categorical variables, as dummy variables;
- For each group formed by categorical predictors,
- The independent variables are not highly correlated with each other.

In short, the Binary Logistic Regression examines if the independent variables have an influence on the probability that the dependent variable takes the value 1 and how strong the influence of the dependent variable is. In our case, the effectiveness of a certain Nudge.

The independent variable of our data set is called 'sustainable item selected'. (1) stands for when the user has chosen the

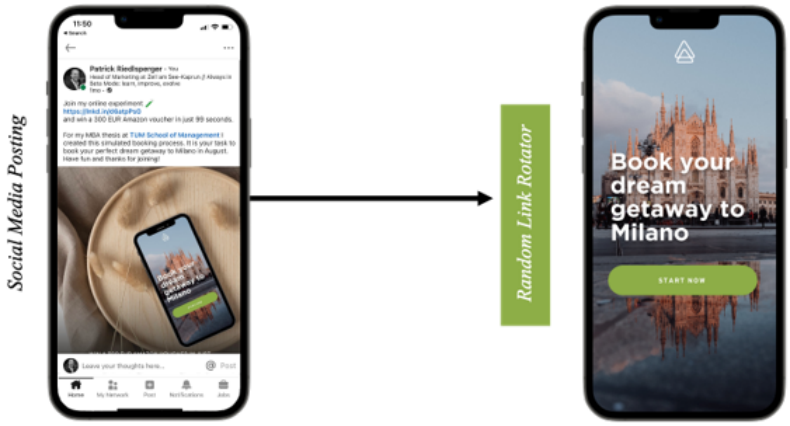


Figure 24: *Generic Social Media Posting on LinkedIn*

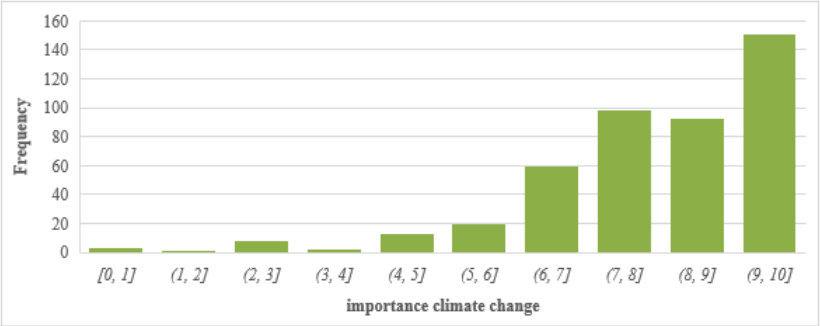


Figure 25: *Frequency of 'importance climate change'*

most sustainable option or positive result, and the (0) stands if the user opted for an alternative choice, or negative result. The dependent variables are different forms of nudging interventions. Therefore, they are categorical and coded as dummy variables for the analysis. For the dependent variables, the "No Nudge" category was set as a reference group. According to Alber (2021) a reference group is a group that we choose to be the reference so that all odds ratios will be a compared to. Therefore, nudging category intervention in the analysis will be compared against the simulated booking process, which did not include any of the interventions. This enables us to measure the effectiveness of Nudges.

The logistic regression analysis is based on Maximum Likelihood Estimation. Similar to a linear regression, an attempt is made to find a function curve that fits the data best. However, in contrast to linear regression analysis, this function is not a straight line but a logistic function. It is "S-shaped", symmetrical and runs asymptotically towards y = 0 and y = 1 – namely positive or negative results of the sample. The model is based on the following formula. $P(y = 1)$ defines the probability that y = 1, e the base of the natural logarithm and z the Logit, which represents a linear regression model of the independent variables (University of Zürich, 2022).

$$P(y = 1) = \frac{1}{1 + e^{-z}}$$

A logistic regression function is not linear and therefore more complex than a linear regression. However, what still applies is if the sign of a regression coefficient is positive, an increase in the relevant independent variable causes an increase in the probability that y = 1. If the sign is negative, this means a decrease in the probability. The connection between an independent variable and the dependent variable can also be interpreted using so-called "odds" in a logistic regression. To calculate the odds, the probability that the event will occur is placed in relation to the non-occurrence of the event. The **Exp(B)** or odds ratios of an independent variable give the change in the relative probability of y = 1. If Exp(B) is one, this results in a multiplication of the relative probability by 1 and thus no change. If the odds ratio > 1, this means an increase in the odds, while an odds ratio < 1 means a decrease in odds (University of Zürich, 2022).

From Exp(B), a relationship can derived, which is very useful for the interpretation of our result:

$$Odds_{after} = Exp(B)\ x\ Odds_{before}$$

$Odds_{before}$ represents the sample of the simulated booking process without choice interventions – our reference group.

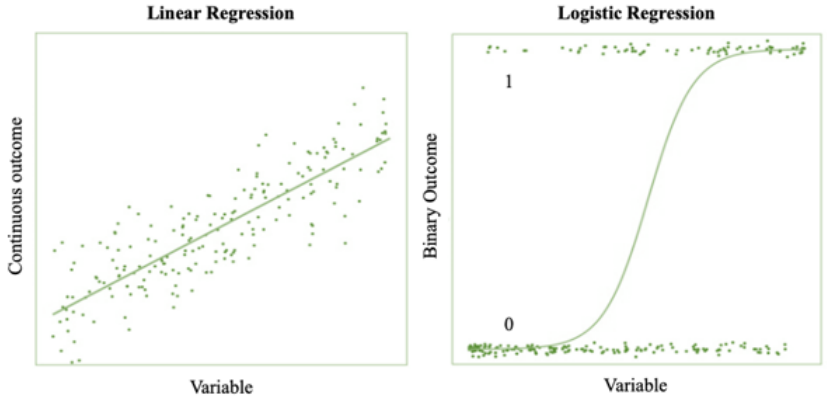


Figure 26: *Linear vs. Logistic regression – Adapted from* Kubben et al. (2019)

Table 11: *Interpretation for regression coefficients and odds ratios. Adapted from University of Zürich (2022)*

B	Exp(B)		P(y=1)
Regression coefficient	Odds ration		
$B > 0$	> 1	increases by the factor Exp (B)	Ascend
$B = 0$	$= 1$	remains the same	Steady
$B > 0$	> 1	decreases by the factor	Decline

$Odds_{after}$ represents the regression coefficient (B) after a nudging intervention has been applied. The higher Exp(B) is the more efficient the form of Nudge. If Exp(B) is > 1, the hypothesis of this research project is confirmed.

The level of statistical significance or probability value is a number describing how likely it is that research data occurred by random chance. A value of ≤ 0.05 is typically defined as statistically significant (McLeod, 2019). For this research project, the significance threshold is also defined in this 0.05 significance threshold. If a value of > 0.05 is obtained for an experiment/nudging test, the result of the Binary Logistic regression is defined as not significant. If the significance value is ≤ 0.05 and the regression coefficient B is > 0, our alternative hypothesis can be confirmed. Based on the significance level, there is a 95% chance that the relationship between the dependent and independent variable is not coincidental.

5. Results

For comparing the different Nudges objectively, the same decision environments have to be evaluated before and after the application of the intervention. Therefore, the 4 different booking questions in the realm of transport, accommodation and restaurants are analyzed individually to gain insights on the effectiveness of nudging people towards the most sustainable travel options. The results are presented in cross-tabs and the described Binary Logistic Regression.

5.1. Arrival

Analyzing the first decision on arrival/mode of transport, the crosstabulation shows that the nudged versions of the simulated booking process had an increased number of decisions towards the most sustainable option (1). The binary logistic regression provided evidence for the effectiveness of nudges as all regression coefficients **(B)** had positive values. When comparing the nudges against each other, it has to be taken into consideration that the G2-6 and the G2-9 experiment uses the disclosure nudge, and the indicated warning and elicitation implementation intentions nudge. The disclosure nudge had an Exp(B) value of 2.068. However, the values for the binary logical regression were not statistically significant. Therefore, no relationship can be inferred based on the regression model.

When the warning or elicitation implementation intentions nudge was placed before the disclosure nudge in experiments (G2-6, G2-9), the Exp(B) values increased. The highest odds for users choosing the most sustainable option were calculated for the G2-6 experiment – the warning nudge. The default nudge G2-1 also increased the odds of choosing the option which emits the least greenhouse gases by a factor auf 2.825. Beyond the experiment G2-5, all results were statistically significant. The first decision of the simulated booking process provided evidence that confirmed our hypothesis. There is evidence that 3 of the 4 Digital nudges lead users towards more sustainable choices.

5.2. Accommodation

For the accommodation decision in the simulated booking process, two Nudges were tested for effectiveness. These were the simplification nudge (G2-2), which used the green hotel label, and the precommitment nudge (G2-7), which used the same label and where participants were additionally asked if protecting the environment was important to

Table 12: *Sustainable Item Selected * Arrival Nudge Crosstabulation*

		Sustainable Item Selected		
	Nudge	***0***	***1***	***Total***
G1	Regular Booking Process	44	6	50
G2-1	Default nudge	35	15	50
G2-5	Disclosure nudge	39	11	50
G2-6	Warning nudge	35	15	50
G2-9	Elicitation implementation intentions nudge	32	18	50

Table 13: *Results Arrival - Binary Logistic Regression*

	B	***S.E.***	***Wald***	***df***	***Sig.***	***Exp(B)***
G1			8.154	4	0.086	
G2-1	1.048	0.537	3.805	1	0.051	**2.852**
G2-5	0.727	0.553	1.726	1	0.189	**2.068**
G2-6	1.417	0.526	7.270	1	0.007	**4.125**
G2-9	1.145	0.534	4.607	1	0.032	**3.143**
Constant	-1.992	0.435	20.960	1	0.000	0.136

them before starting the booking process. The two options were compared against how many users chose the same hotel without intervention.

Out of the 50 respondents for G2-7, 44 testers selected that protecting the environment is important to them. 28 of the testers took the most sustainable accommodation option in the booking process afterwards. Comparatively in the regular booking process, only 17 users chose the same hotel offer and only 20 users selected that option in the version with the simplification intervention.

For our G2-2 experiment – the simplification nudge - the binary logistic regression calculated only a slight increase of odds Exp(B) of 1.078. These results were not statistically significant. However, when users pre-committed to protect our environment and were offered the sustainable option afterward, Exp(B) increased to 2.471 with statistically significant results.

5.3. Transport

Analyzing the decision on users choosing either public transport or an alternative option, there were only minor differences between the nudged version and the regular booking process.

The majority of users (69%) chose to use public transport during their stay in Milano. For the regular booking process (G1), 34 people chose the sustainable option. The G2-3 experiment, where the size of the interaction element was increased, had 35 users choosing public transport.

Running binary logistic regression, the relationship between the dependent variable, sustainable item selected, and the independent nudge was not statistically significant. Therefore, no relationship was proven between the Digital nudge and the selection of the most sustainable item. despite many users choosing the climate-friendly public transport option in the experiment G1 and G2-3. This could be related to the convenience of the actual service as public transport is inexpensive, fast and punctual in many European metropolitan areas.

5.4. Meals

The final analysis of the nudging experiments relates to the selection of meals. Hereby, the experiment measured how many users opted for the most sustainable, vegan meal option.

In the regular booking process, only 3 users opted for the vegan option. When the reminder nudge was used in the experiment G2-8, 7 research participants switched their diet option, which resulted in a total of 10 people selecting the vegan choice. The social norm nudge - which indicated that 64% of other guests chose to eat vegan dishes during their stay - led 11 users to make the same decision and reduce the overall greenhouse gas emission of their stay.

For both options, regression coefficients **(B)** were positive and results were statistically significant. The higher Exp(B) was calculated in the regression for the G2-4 experiment – the social norm nudge. Compared to the regular booking process, odds increased by 4.419. The reminder nudge also led a notable increase in odds Exp(B) of 3.917, based on the effects of the intervention.

Comparing the regular simulated booking process to the nudged versions, this study observed the intervention led more users to choose the most sustainable items (1). Overall, the digital nudges in the observed experiments G2-1 to G2-9 had a positive regression coefficient (B) and thereby odds ratio Exp(B) > 0. Of the 9 experiments which used digital nudging interventions, 6 produced statistically significant results. Their odds ratios were between 2.471 and 4.419. Based on this observation, the alternative hypothesis (H1) can be confirmed.

> ✓ *Digital nudges lead to the booking of the most sustainable travel offer*

Table 14: *Sustainable Item Selected * Accommodation Nudge Crosstabulation*

		Sustainable Item Selected		
	Nudge	***0***	***1***	***Total***
G1	Regular Booking Process	33	17	50
G2-2	Simplification Nudge	36	20	56
G2-7	Precommitment Nudge	22	28	50

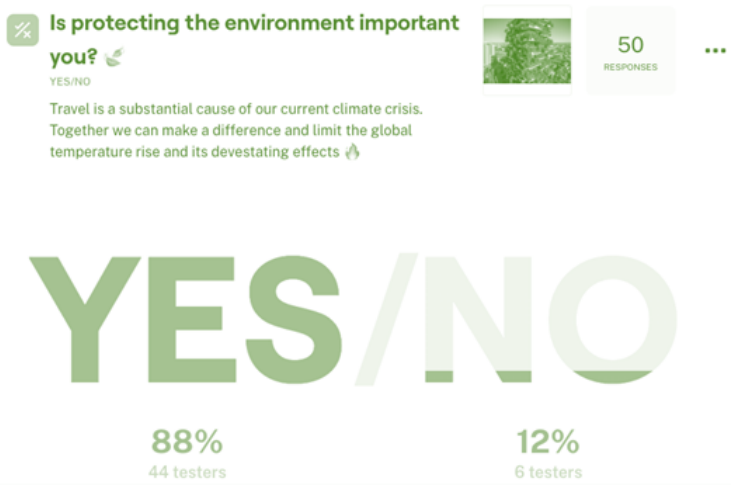


Figure 27: *Precommitment Nudge Results (Screenshot)*

Table 15: *Results Accommodation - Binary Logistic Regression*

		B	***S.E.***	***Wald***	***df***	***Sig.***	***Exp(B)***
	G1			6.120	2	0.047	
	G2-2	0.076	0.409	0.034	1	0.853	**1.078**
	G2-7	0.904	0.413	4.804	1	0.028	**2.471**
	Constant	-0.663	0.299	4.936	1	0.026	0.515

Table 16: *Sustainable Item Selected * Transport Nudge Crosstabulation*

		Sustainable Item Selected		
	Nudge	***0***	***1***	***Total***
G1	Regular Booking Process	16	34	50
G2-3	Increase in ease and convenience nudge	15	35	50

Table 17: *Results Transport - Binary Logistic Regression*

		B	***S.E.***	***Wald***	***df***	***Sig.***	***Exp(B)***
	G2-3	0.094	0.433	0.047	1	0.829	**1.098**
	Constant	-0.663	0.299	4.936	1	0.026	0.515

Table 18: *Sustainable Item Selected * Meals nudge Crosstabulation*

		Sustainable Item Selected		
	Nudge	***0***	***1***	***Total***
G1	Regular Booking Process	47	3	50
G2-4	Social Norm Nudge	39	11	50
G2-8	Reminder Nudge	40	10	50

Table 19: *Results Meals - Binary Logistic Regression*

	B	*S.E.*	*Wald*	*df*	*Sig.*	*Exp(B)*
G1			4.973	2	0.083	
G2-4	1.486	0.686	4.686	1	0.030	4.419
G2-8	1.365	0.693	3.886	1	0.049	3.917
Constant	-2.752	0.595	21.350	1	0.000	0.064

The most effective Digital nudge was observed at the G2-4 experiment (4.419) followed by the G2-6 experiment (4.125). Choice architects can increase the odds of users choosing the most sustainable option by more than 4 times. Furthermore, the experiment G2-8 provided evidence that nudging users by reminding them to take a more sustainable choice, nearly quadruples (3.917) the amount of people opting for the decision which emits the least greenhouse gases. Additionally, the elicitation implementation intentions nudge (G2-9) is an effective way to reduce emissions for travel bookings. It provides user interface designers with the ability to more than triple (3.142) the selection of an indicated sustainable product category. The default nudge (G2-1) also nearly tripled (2.852) the chances of users to select the most sustainable arrival option in the experiment. The precommitment nudge was the statistically valid result, with the merest degree of effectiveness – though it increased odds by 2.471.

6. Conclusion

Individual consumption decisions we make every day have caused our current climate crisis. Too often, we operate on System 1, which leads us to short-sighted and self-centered decisions made at the expense of our climate and planet. This includes our decision-making when we are booking a holiday on digital channels. This thesis has been able to demonstrate that nudges can help us overcome this issue. As Sunstein and Thaler (2021) indicated, the experiments proved that in a travel context, environmentally-friendly nudges can help reduce air pollution and the emission of greenhouse gases. This doesn't mean that they can totally resolve the challenge of global warming and travel-related emissions. Nudges are not the solution, but they can be one of the many small steps we take to reach a net zero emission economy by 2050. Beyond technological innovation, fiscal and regulatory measures, nudges are a simple, cost-effective way to reduce the level of greenhouse gas emissions in the travel industry.

Research participants were very aware of the issue of climate change with 76% rating it in an 8-10 range. However, when analyzing those participants in detail, the study observed only 38% of them chose the most sustainable option in the booking process. The experiments found the same phenomenon that Hornsey et al. (2016), Jacobsen (2011) and Mazar et al. (2020) discovered in their research projects. Our society is not lacking environmentally-friendly attitudes or intentions. It is lacking in transforming those into actions and results. Nudges can help us achieving that. This research did not calculate the overall greenhouse gas emissions which would be saved by implementing our green digital nudges. Although this would be an interesting future experiment, the increase in odds or Exp(B) ratios provides evidence that overall emissions can be drastically lowered by choice interventions. People still want to and should travel in the future. It can increase health, enhance creativity, and exposes us to new cultures and environments. The overall greenhouse emission can be lowered by green nudges along the customer journey.

Regarding the effectiveness of the different nudging interventions, this research confirms the literature by Kosters and van der Heijden (2015) and Sunstein and Thaler (2021) that the virtue of nudges is context-specific. While Hummel and Maedche (2019) tried to generalize results for different nudging research projects, this thesis did not confirm their findings on the effectiveness of nudges. For a travel-specific context, default nudges were not the most efficient form of choice intervention. While the nudge ten folded the purchases of a green energy plan according to Ebeling and Berger (2015), this travel-specific experiment led to a near tripling of sales of the sustainable product (2.852). Based on the results of this thesis, choice architects should use social norm nudges if they aim to only implement one specific intervention. However, hybrid nudges which use two or more interventions have also proven to be successful in the experiments G2-6 to G2-8. They remind users to act sustainably, and received statistically significant results contrary to the implementation of the interventions as a singular digital nudge in the G2-2 and G2-5 experiment.

Discussing the limitations of this research, the simulation of the booking process could have led to the Hawthorne Effect. This effect may have reduced the validity of the results. According to Koch (2011) the effect describes the phenomenon of test subjects changing their output on a test in response to being observed, as some Hawthorne employees did when they knew they were part of the study. Researchers today use randomized control groups in experiments to weed out this bias in studies. Although this control group has been created in the experiment with the regular booking process, the nudged versions could have generated more users to choose the socially desirable behavior i.e., the selection of the most sustainable option. Especially when warning users, or after a precommitment to take climate action, or the elicitation of implementation intentions, some users could have made biased decisions in the booking process. Furthermore, it is unclear for all experiments including the regular booking process if users would have behaved in the exact same way

Table 20: *Experiment Overview - Effectiveness of Digital nudges*

Experiment / Nudge		*Exp(B)*	*Significant*
G1	Regular Booking Process	-	-
G2-4	Social norm nudge	**4.419**	✓
G2-6	Warning nudge	**4.125**	✓
G2-8	Reminder nudge	**3.917**	✓
G2-9	Elicitation implementation intentions nudge	**3.143**	✓
G2-1	Default nudge	**2.852**	✓
G2-7	Precommitment nudge	**2.471**	✓
G2-5	Disclosure nudge	**2.068**	
G2-3	Increase in ease and convenience nudge	**1.098**	
G2-2	Simplification nudge	**1.078**	

in an uncontrolled field experiment.

Future research on the topic could include the application of the designed digital nudges into an actual e-commerce environment to measure their effectiveness on the booking of more sustainable products. In these experiments, researchers could gain larger samples and include other influencing factors such as prices in future studies. As the process of this research project has been a well-designed, but simulated environment, there is uncertainty around whether testers would have behaved the same way if an actual payment follows after the completion of the booking process. It would also be interesting to observe if people are willing to pay higher prices for more sustainable products, if they are nudged for the suitable options. Paying higher prices can also include the offsetting of carbon emissions for services, where fossil fuels are still used. As discussed, one important factor beyond the digital application of nudges is the convenience of the physical travel services and products.

Nudges can guide people in their decision making, but they won't let people choose options for decisions against which people have an aversion based on their prior experience. Therefore, sustainable transfer options have to be convenient and punctual, green hotels have to provide the same excellent services as their competitors, and vegan food has to be tasty. This research project has proven that decarbonizing travel decisions by using nudges is an effective strategy and one of the many steps we can perform to fight our current, global climate crisis. Nudges enable us to enhance the decision-making of every traveler to a more sustainable outcome. The thesis should encourage accommodation and mobility providers, online travel agents, tour operators and other participants across the value chain of travel to implement those nudges in their digital booking solutions. When it comes to the conservation of greenhouse gases, every individual decision counts and adds up to reduce global warming. Choice interventions in a digital travel booking process can make it more than 4 times likely that users take the most sustainable option. Let's nudge together for a better, more sustainable future.

References

Alber, S. (2021). Logistic Regression. *UC Davis Health*. https://health.ucdavis.edu/media-resources/ctsc/documents/pdfs/logistic-regression-intro-2021.pdf

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, *95*(9-10), 1082–1095.

Allen, M. (2017). *The SAGE Encyclopedia of Communication Research Methods*. SAGE Publications, Inc. https://doi.org/10.4135/9781483381411

Amram, M., & Kulatilaka, N. (2009). The Invisible Green Hand: How Individual Decisions and Markets Can Reduce Greenhouse Gas Emissions. *California Management Review*, *51*. https://doi.org/10.2307/41166486

Andersson, H. (2019). *Designing Digital Nudges for Sustainable Travel Decisions* [Doctoral dissertation, Umea University].

Arora, V. (2021). Mobile Commerce's Growing Opportunities for Travel: Latest Skift Research. *Skift Research*. https://skift.com/2021/10/27/mobile-commerces-growing-opportunities-for-travel-latest-skift-research/

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. *Organizational Influence Processes*, *58*, 295–303.

Berger, S. (2015). The Power of the Nudge to Change Our Energy Future: A simple change, based on psychology, that increases use of renewable energy. https://www.scientificamerican.com/article/the-power-of-the-nudge-to-change-our-energy-future/

Beshears, J., & Kosowsky, H. (2020). Nudging: Progress to date and future directions. *Organizational Behavior and Human Decision Processes*, *161*, 3–19. https://doi.org/10.1016/j.obhdp.2020.09.001

Booking.com. (n.d.). Sustainability solutions - Booking.com Partner Hub. https://partner.booking.com/en-gb/solutions/advice/sustainability-solutions

Budiu, R., & Moran, K. (2021). How Many Participants for Quantitative Usability Studies: A Summary of Sample-Size Recommendations. *Nielsen Norman Group*. https://www.nngroup.com/articles/summary-quant-sample-sizes/

Cantor, J., Torres, A., Abrams, C., & Elbel, B. (2015). Five Years Later: Awareness Of New York City's Calorie Labels Declined, With No Changes In Calories Purchased. *Health Affairs (Project Hope)*, *34*(11), 1893–1900. https://doi.org/10.1377/hlthaff.2015.0623

Chambliss, D. F., & Schutt, R. K. (2016). *Making Sense of the Social World: Methods of Investigation* (Fifth). SAGE.

Christensen, C. (2020, January 25). Interview by C. Sakar.

Cooper, E. J. (2017). To Nudge or Not to Nudge: Promoting Environmentally Beneficial Behaviors.

Creswell, J. (2009). *Research Design: Qualitative, Quantitative, and Mixed-Method Approaches*. SAGE.

DaSilva, B., Dhar, J., Rafiq, S., & Young, D. (2022). Nudging Consumers Toward Sustainability. *BCG Henderson Institute*. https://www.bcg.com/publications/2022/nudging-consumers-to-make-sustainable-choices

Deutsche Bahn. (2022). Deutsche Bahn Website. https://www.bahn.de/

Dhar, J., Bailey, A., Mingardon, S., & Tankersley, J. (2017). The Persuasive Power of the Digital Nudge. *Boston Consulting Group*. https://www.bcg.com/publications/2017/people-organization-operations-persuasive-power-digital-nudge

Easterby-Smith, M., Thorpe, R., & Jackson, P. R. (2015). *Management and business research* (5th). SAGE.

Ebeling, F., & Berger, S. (2015). Domestic uptake of green energy promoted by opt-out tariffs. *Nature Climate Change*, *5*. https://doi.org/10.1038/nclimate2681

Engel, H., Walter, D., & Olynyk, G. (2022). Failure is not an option: Increasing the chances of achieving net zero. *McKinsey & Company*. https://www.mckinsey.com/business-functions/sustainability/our-insights/failure-is-not-an-option-increasing-the-chances-of-achieving-net-zero

European Environment Agency. (2021). Climate change is one of the biggest challenges of our times. https://www.eea.europa.eu/themes/climate/climate-change-is-one-of

Fischetti, M. (2021). There's Still Time to Fix Climate—About 11 Years: Aggressive policies, enacted now, can extend the deadline and prevent the worst catastrophes. https://www.scientificamerican.com/article/theres-still-time-to-fix-climate-about-11-years/

Flixbus. (2022). Flixbus Website. https://www.flixbus.de/

Fountain, H. (2022). Why Europe Is Becoming a Heat Wave Hot Spot. *New York Times*. https://www.nytimes.com/2022/07/19/climate/europe-heat-wave-science.html

Franklin, M., Folke, T., & Ruggeri, K. (2019). Optimising nudges and boosts for financial decisions under uncertainty. *Palgrave Communications*, *5*(1), 113. https://doi.org/10.1057/s41599-019-0321-y

Friedman, M., & Friedman, R. D. (1990). *Free to choose: A personal statement* (1st Harvest/HBJ ed.). Harcourt Brace Jovanovich.

Fritz, M., & Berger, P. D. (2015). *Improving the User Experience Through Practical Data Analytics*. Morgan Kaufmann.

Frölicher, T., Winton, M., & Sarmiento, J. (2013). Continued global warming after CO2 emissions stoppage. *Nature Climate Change*, *4*. https://doi.org/10.1038/NCLIMATE2060

Fung, K. (2017, July 18). Interview by A. Gallo [Harvard Business Review].

García, J. M., & Vila, J. (2020). Financial literacy is not enough: The role of nudging toward adequate long-term saving behavior. *Journal of Business Research*, *112*, 472–477. https://doi.org/10.1016/j.jbusres.2020.01.061

Gates, B. (2021, March 1). Interview by A. Ignatius [Harvard Business Review].

Gilbert, E. (2021). Being Spock (and not being Homer). *Blue Chip Digital*, (78), 54–55.

Gimpel, H., Heger, S., Kasper, J., & Schäfer, R. (2020). The Power of Related Articles – Improving Fake News Detection on Social Media Platforms. https://doi.org/10.24251/HICSS.2020.743

Gollwitzer, P. (1999). Implementation Intentions: Strong Effects of Simple Plans. *American Psychologist*, *54*, 493–503. https://doi.org/10.1037/0003-066X.54.7.493

Google Flights. (2022). Google Flight Search. https://www.google.com/flights

Google Maps. (2022). Google Maps Search. https://www.google.com/maps

Gore, A. (2006). An Inconvenient Truth.

Hallsworth, M., & Kirkman, E. (2020). *Behavioral insights*. The MIT Press.

Hanks, A. S., Just, D. R., Smith, L. E., & Wansink, B. (2012). Healthy convenience: nudging students toward healthier choices in the lunchroom. *Journal of Public Health*, *34*(3), 370–376. https://doi.org/10.1093/pubmed/fds003

Hansen, P. G. (2016). The Definition of Nudge and Libertarian Paternalism: Does the Hand Fit the Glove? *European Journal of Risk Regulation*, *7*(1), 155–174. https://doi.org/10.1017/S1867299X00005468

Harrison, G. W., & List, J. A. (2004). Field Experiments. *Journal of Economic Literature*, *42*(4), 1009–1055. http://www.jstor.org/stable/3594915

Heitmann, M., Goldstein, D., Johnson, E., & Herrmann, A. (2008). Nudge Your Customers Toward Better Choices. *Harvard Business Review*, *86*, 99–106.

Hornsey, M., Harris, E., Bain, P., & Fielding, K. (2016). Meta-analyses of the determinants and outcomes of belief in climate change. *Nature Climate Change*, *6*. https://doi.org/10.1038/nclimate2943

House of Lords. (2011). Behaviour Change. https://publications.parliament.uk/pa/ld201012/ldselect/ldsctech/179/179.pdf

Hummel, D., & Maedche, A. (2019). How effective is nudging? A quantitative review on the effect sizes and limits of empirical nudging studies. *Journal of Behavioral and Experimental Economics*, *80*, 47–58. https://doi.org/10.1016/j.socec.2019.03.005

IPCC. (2021). Climate Change 2021: The Physical Science Basis: The Working Group I contribution to the Sixth Assessment Report. https://www.ipcc.ch/report/ar6/wg1/

Jackson, R. L., Drummond, D. K., & Camara, S. (2007). What Is Qualitative Research? *Qualitative Research Reports in Communication*, *8*(1), 21–28. https://doi.org/10.1080/17459430701617879

Jacobsen, G. D. (2011). The Al Gore effect: An Inconvenient Truth and voluntary carbon offsets. *Journal of Environmental Economics and Management*, *61*(1), 67–78. https://doi.org/10.1016/j.jeem.2010.08.002

Jesse, M., Jannach, D., & Gula, B. (2021). Digital Nudging for Online Food Choices. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.729589

Kahneman, D. (2013). *Thinking, fast and slow* (1st pbk. ed.). Farrar Straus and Giroux.

Karlsen, R., & Andersen, A. (2019). Recommendations with a Nudge. *Technologies*, *7*, 45. https://doi.org/10.3390/technologies7020045

King, R., Churchill, E. F., & Tan, C. (2017). *Designing with data: Improving the user experience with A/B testing* (First edition). O'Reilly.

Koch, K. (2011). Rethinking work, beyond the paycheck. https://news.harvard.edu/gazette/story/2011/12/rethinking-work-beyond-the-paycheck

Kohavi, R., & Thomke, S. (2017). The Surprising Power of Online Experiments. *Harvard Business Review*. https://hbr.org/2017/09/the-surprising-power-of-online-experiments

Kosters, M., & van der Heijden, J. (2015). From mechanism to virtue: Evaluating Nudge theory. *Evaluation*, *21*(3), 276–291. https://doi.org/10.1177/1356389015590218

Kubben, P., Dumontier, M., & Dekker, A. (2019). *Fundamentals of clinical data science*. Springer.

Lade, G., Jessoe, K., Loge, F., & Spang, E. (2020). Spillovers from Behavioral Interventions: Experimental Evidence from Water and Energy Use. https://doi.org/10.7910/DVN/THWDLR

Lehner, M., Mont, O., & Heiskanen, E. (2016). Nudging – A promising tool for sustainable consumption behaviour? *Journal of Cleaner Production*, *134*, 166–177. https://doi.org/10.1016/j.jclepro.2015.11.086

Lenzen, M., Sun, Y.-Y., Faturay, F., Ting, Y.-P., Geschke, A., & Malik, A. (2018). The carbon footprint of global tourism. *Nature Climate Change*, *8*(6), 522–528. https://doi.org/10.1038/s41558-018-0141-x

Leonard, T. C. (2008). Richard H. Thaler, Cass R. Sunstein, Nudge: Improving decisions about health, wealth, and happiness. *Constitutional Political Economy*, *19*(4), 356–360. https://doi.org/10.1007/s10602-008-9056-2

Levy, J. (2015). *UX strategy: How to devise innovative digital products that people want* (First edition). O'Reilly Media Inc.

Mazar, A., Tomaino, G., Carmon, Z., & Wood, W. (2020). Habits to Save our Habitat: Using the Psychology of Habits to Promote Sustainability. https://doi.org/10.31234/osf.io/kwdh9

McLeod, S. (2019). What a p-value tells you about statistical significance. https://www.simplypsychology.org/p-value.html

Mertens, S., Herberz, M., Hahnel, U. J. J., & Brosch, T. (2022). The effectiveness of nudging: A meta-analysis of choice architecture interventions across behavioral domains. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(1). https://doi.org/10.1073/pnas.2107346118

Milkman, K. (2022). A 680,000-person megastudy of nudges to encourage vaccination in pharmacies. *Proceedings of the National Academy of Sciences*, *119*(6), e2115126119. https://doi.org/10.1073/pnas.2115126119

Mirsch, T., Jung, R., Rieder, A., & Lehrer, C. (2018). Mit Digital Nudging Nutzererlebnisse verbessern und den Unternehmenserfolg steigern. *Marketing ZFP - Journal of Research and Management*, *30*, 12–18. https://doi.org/10.15358/0935-0381-2018-5-12

Mont, O., Lehner, M., & Heiskanen, E. (2014). Nudging: A tool for sustainable behaviour? *Naturvårdsverket*.

Nahmias, Y. (2019). Privacy Preserving Social Norm Nudges. *Michigan Telecommunications and Technology Law Review*, *26*. https://repository.law.umich.edu/mtlr/vol26/iss1/3

Nielsen. (2007). Climate Change & Influential Spokespeople.

Nielsen, A. S. E. (2016). Nudging and pro-environmental behaviour. *Nordic Council of Ministers*.

Nijhuis, M. (2020). *Nudging tourists: The application of nudging within tourism to achieve sustainable travelling* [Master's thesis, Universiteit Twente].

Nudges that don't nudge. (2020). *Nature Human Behaviour*, *4*(2), 121. https://doi.org/10.1038/s41562-020-0832-y

Oreopoulos, P., & Petronijevic, U. (2019). The Remarkable Unresponsiveness of College Students to Nudging and What We Can Learn from it. *DecisionSciRN: Decision Making in Education (Topic)*.

Paris Agreement. (2016). https://unfccc.int/sites/default/files/english_paris_agreement.pdf

Pateil, N. (2022). A/B Testing: Definition, How it Works, Examples, & Tools. https://neilpatel.com/blog/ab-testing-introduction/

Pereira, F. (2019). *Digital Nudge: The hidden forces behind the 35,000 decisions we make every day*. EDITORA CAROLI.

Porter, M. E., & van der Linde, C. (1995). Green and Competitive: Ending the Stalemate.

Rauner, M. (2015). Die Fliege im Klo - und die Stupser der Kanzlerin. https://www.zeit.de/zeit-wissen/2014/06/nudging-politik-verhaltensforschung-psychologie

Ries, E. (2011). *The lean startup: How today's entrepreneurs use continuous innovation to create radically successful businesses*. Crown Business.

Satouri, N. (2019). AB-Testing Trivago. *Github*.

Scarborough, P., Appleby, P., Mizdrak, A., Briggs, A., Travis, R., Bradbury, K., & Key, T. (2014). Dietary greenhouse gas emissions of meat-eaters, fish-eaters, vegetarians and vegans in the UK. *Climatic Change*, *125*. https://doi.org/10.1007/s10584-014-1169-1

Schellnhuber, H. J. (2021). Interview by U. Sommersguter. https://www.kleinezeitung.at/wirtschaft/5964710/Klimaforscher-Schellnhuber_Klimawandel-aehnlicher-Notfall-wie

Schmidt, A. (2017). The Power to Nudge. *American Political Science Review*, *111*(2), 404–417. https://doi.org/10.1017/S0003055417000028

Schmidt, A., & Engelen, B. (2020). The ethics of nudging: An overview. *Philosophy Compass*, *15*(4), e12658. https://doi.org/10.1111/phc3.12658

Schneider, C., Weinmann, M., & vom Brocke, J. (2018). Digital Nudging: Guiding Online User Choices through Interface Design. *Communications of the ACM*, *61*, 67–73. https://doi.org/10.1145/3213765

Siipi, H., & Koi, P. (2022). The Ethics of Climate Nudges: Central Issues for Applying Choice Architecture Interventions to Climate Policy. *European Journal of Risk Regulation*, *13*(2), 218–235. https://doi.org/10.1017/err.2021.49

Siroker, D., & Koomen, P. (2015). *A/B testing: The most powerful way to turn clicks into customers*. John Wiley & Sons, Inc.

Sperling, J., & Güntner, A. (2017). How to nudge your way to better performance. *McKinsey & Company*. https://www.mckinsey.com/business-functions/people-and-organizational-performance/our-insights/the-organization-blog/how-to-nudge-your-way-to-better-performance

Spiliakos, A. (2017). Nudging Change: Why We Do What We Do. *Harvard Business School*. https://online.hbs.edu/blog/post/nudging-change-why-we-do-what-we-do

Sunstein, C. (2014). The Ethics of Nudging. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2526341

Sunstein, C. (2015). Nudging and Choice Architecture: Ethical Considerations. *Harvard John M. Olin Discussion Paper Series, Discussion Paper No. 809*.

Sunstein, C. (2016). People Prefer System 2 Nudges (Kind Of). *Duke Law Journal*, *66*(1), 121–168. http://www.jstor.org/stable/44155309

Sunstein, C. (2017a). *Human Agency and Behavioral Economics: Nudging Fast and Slow* (1st). Springer International Publishing; Imprint: Palgrave Macmillan.

Sunstein, C. (2017b). Nudges that fail. *Behavioural Public Policy*, *1*(1), 4–25. https://doi.org/10.1017/bpp.2016.3

Sunstein, C. (2019). Nudging: a very short guide. *Business Economics*, *54*(2), 127–129. https://doi.org/10.1057/s11369-018-00104-5

Sunstein, C., & Thaler, R. H. (2003). Libertarian Paternalism. *The American Economic Review*, *93*(2), 175–179. http://www.jstor.org/stable/3132220

Sunstein, C., & Thaler, R. H. (2021). *Nudge: Improving decisions about money, health, and the environment* (Final edition). Penguin Books, an imprint of Penguin Random House LLC.

Szaszi, B., Palinkas, A., Palfi, B., Szollosi, A., & Aczel, B. (2018). A Systematic Scoping Review of the Choice Architecture Movement: Toward Understanding When and Why Nudges Work. *Journal of Behavioral Decision Making*, *31*(3), 355–366. https://doi.org/10.1002/bdm.2035

Tett, G. (2021, September). Could behavioural nudges help us tackle the climate crisis? *FT Magazine*. https://www.ft.com/content/b6d36790-b171-417c-b593-b26900df3cd5

Thaler, R. H. (2009). Do you need a nudge? Yale Insights - Management in Practice. https://insights.som.yale.edu/insights/do-you-need-nudge

Thaler, R. H. (2011, November 1). Interview by D. Lovallo & W. Allen. *McKinsey Quarterly*.

Thaler, R. H. (2015). The Power of Nudges, for Good and Bad. https://www.nytimes.com/2015/11/01/upshot/the-power-of-nudges-for-good-and-bad.html

Thaler, R. H., & Benartzi, S. (2004). Save More Tomorrow™: Using Behavioral Economics to Increase Employee Saving. *Journal of Political Economy*, *112*(S1), S164–S187. https://doi.org/10.1086/380085

Thaler, R. H., & Sunstein, C. (2021, August 6). Interview by R. Fusaro & J. Sperling-Magro.

Thorun, C., Diels, J., Vetter, M., Reisch, L., Bernauer, M., Micklitz, H.-W., Rosenow, J., Forster, D., Sunstein, C., & Purnhagen, K. (2017). Nudge-Ansätze beim nachhaltigen Konsum: Ermittlung und Entwicklung von Maßnahmen zum „Anstoßen" nachhaltiger Konsummuster.

Tollefson, J. (2021). IPCC climate report: Earth is warmer than it's been in 125,000 years. *Nature*, *596*(7871), 171–172. https://doi.org/10.1038/d41586-021-02179-1

Tucker, W., & Thaler, R. H. (2013). Smarter Information, Smarter Consumers.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, *185*(4157), 1124–1131. http://www.jstor.org/stable/1738360

Umweltbundesamt. (2021). Vergleich der durchschnittlichen Emissionen einzelner Verkehrsmittel im Personenverkehr in Deutschland. https://www.umweltbundesamt.de/bild/vergleich-der-durchschnittlichen-emissionen-0

United Nations. (2022). What Is Climate Change? https://www.un.org/en/climatechange/what-is-climate-change

University of Zürich. (2022). Logistische Regressionsanalyse. https://www.methodenberatung.uzh.ch/de/datenanalyse_spss/zusammenhaenge/lreg.html

UNWTO. (2019). Transport-related CO2 Emissions of the Tourism Sector – Modelling Results. https://www.e-unwto.org/doi/book/10.18111/9789284416660

U.S. Food and Drug Administration. (2019, August 15). FDA proposes new required health warnings with color images for cigarette packages and advertisements to promote greater public understanding of negative health consequences of smoking. https://www.fda.gov/news-events/press-announcements/fda-proposes-new-required-health-warnings-color-images-cigarette-packages-and-advertisements-promote

Vallee, A. (2016). What is A/B testing & What is it used for? *Harvard Business School*. https://online.hbs.edu/blog/post/what-is-ab-testing

Visit Iceland. (2022). The Icelandic Pledge. https://pledge.visiticeland.com/

Vugt, M., Griskevicius, V., & Schultz, P. (2014). Naturally Green: Harnessing Stone Age Psychological Biases to Foster Environmental Behavior. *Social Issues and Policy Review*, *8*. https://doi.org/10.1111/sipr.12000

Wall, T., Cooper, C. L., & Brough, P. (2021). *The SAGE handbook of organizational wellbeing* (1st). SAGE.

Wang, U. (2014). New Technology Tools Aim to Reduce Water Use. *The Wall Street Journal*. https://www.wsj.com/articles/SB10001424052702303647204579543781303147754

Weinmann, M., Schneider, C., & vom Brocke, J. (2016). Digital Nudging. *Business & Information Systems Engineering*, *58*(6), 433–436. https://doi.org/10.1007/s12599-016-0453-1

Wiggers, K. (2022). Maze raises $40M to help facilitate software product research. *TechCrunch*. https://techcrunch.com/2022/06/09/2331513/

Wilkinson, M. (2013). Nudges manipulate, except when they don't. *London School of Ecnomics*. https://blogs.lse.ac.uk/politicsandpolicy/nudges-manipulate-except-when-they-dont/

Williams, M., Wiggins, R. D., & Vogt, W. P. (2021). *Beginning quantitative research. The Sage quantitative research kit.* Sage Publications.

Witzenleiter, M. (Ed.). (2021). *Quick Guide A/B Testing: Wie Sie Ihr Website- und E-Commerce-Testing erfolgreich auf- und umsetzen*. Springer Fachmedien Wiesbaden.

Woetzel, J., Pinner, D., Samandari, H., Engel, H., Krishnan, M., & Boland, B. (2020). Climate risk and response: Physical hazards and socioeconomic impacts. https://www.mckinsey.com/business-functions/sustainability/our-insights/climate-risk-and-response-physical-hazards-and-socioeconomic-impacts

World Meteorological Organization. (2022, May 9). WMO update: 50:50 chance of global temperature temporarily reaching 1.5°C threshold in next five years. https://wmo.int/news/media-centre/wmo-update-5050-chance-of-global-temperature-temporarily-reaching-15degc-threshold-next-five-years

World Tourism Organization. (2008). Climate change and tourism - Responding to global challenges. *UNWTO*.

World Travel & Tourism Council. (2021). A Net Zero Roadmap for Travel & Tourism: Proposing a new target Framework for the Travel & Tourism Sector.

World Travel & Tourism Council. (2022). Travel & Tourism Economic Impact 2022. https://wttc.org/Research/Economic-Impact

Wynes, S., & Nicholas, K. (2017). The climate mitigation gap: Education and government recommendations miss the most effective individual actions. *Environmental Research Letters*, *12*, 74024. https://doi.org/10.1088/1748-9326/aa7541

Yale. (2022). Experimental Design. http://www.stat.yale.edu/Courses/1997-98/101/expdes.htm

Zimmermann, S., Hein, A., Schulz, T., Gewald, H., & Krcmar, H. (2021). Digital Nudging Toward Pro-Environmental Behavior: A Literature Review Completed Research Paper.

Junior Management Science 9(1) (2024) 1211-1240

Junior Management Science

www.jums.academy
ISSN: 2942-1861

Depolarizing Innovation: Dynamic Policy Implications for Entrepreneurial Ecosystems in Second-Tier European Regions

Jan Keim

Trinity College Dublin, The University of Dublin

Abstract

Entrepreneurial ecosystem (EE) research has mainly focused on metropolitan regions and neglected second-tier (European) regions. I use a comparative case study approach with a focus on regional public policy to analyze two second-tier European regions: Uppsala and Galway. The results show that EEs can emerge as a by-product of attracting foreign direct investment or investment in higher education and research. In both cases, the R&D activities of multinational enterprises (MNEs) and universities contribute to the emergence and growth of EEs by enabling the creation of spin-offs. Given the limited resources in second-tier regions, EE initially focus on specific industry clusters to maximize resource efficiency. Later diversification increases ecosystem resilience and mitigates cluster risks. However, limited access to growth capital in second-tier EEs leads to increased acquisition activity by MNEs or the relocation of high-growth ventures to metropolitan areas. Policy measures that support second-tier regions' efforts to create local EEs initially focus on promoting R&D, knowledge spillovers, and research commercialization, later include the creation of supportive infrastructure, and finally enable the attraction of growth capital to the region.

Keywords: economic geography; entrepreneurial ecosystems; public policy; second-tier regions; spatial context

1. Introduction

This article focuses on the evolution of entrepreneurial ecosystems in second-tier European regions. Section 1 introduces the context of the article and defines the research question that forms the basis for all subsequent sections.

1.1. Context

Entrepreneurial ecosystems (EEs) have recently gained considerable attention in the academic literature and among policymakers (Roundy, 2017). The interest is motivated by the argument that entrepreneurship can drive economic development, employment levels, and productivity growth (Isenberg, 2010; Szerb et al., 2015). Many studies on EEs tend to focus on the national level (Audretsch et al., 2015). Yet, the regional spatial context of entrepreneurship within a country influences the outcome of entrepreneurial activities due to different social, institutional, and economic factors (Müller & Korsgaard, 2018; Roundy, 2017). Indeed, Glaeser et al. (2011) note that regional economic development can differ greatly from national economic development. Therefore, regional differences that determine the context of entrepreneurial activity should be considered when analyzing EEs (Audretsch et al., 2015).

Certain cities and regions around the world have become hubs for startups and innovation-driven companies. Most of these hubs are located in larger "superstar" metropolitan areas (e.g., Kemeny and Storper, 2020; Atkinson et al., 2019). But when it comes to entrepreneurial activity, smaller cities and regions should not be neglected. On the one hand, growing territorial inequality due to a polarization of innovation

I would like to thank my supervisor Dr. Giulio Buciuni from the Faculty of Arts, Humanities and Social Sciences at Trinity College Dublin for his continuous support and academic and personal guidance. He was always available if I had a question about my research or writing. Even though the Covid-19 pandemic posed an unexpected challenge to my research, Giulio supported me in every way. I am wholeheartedly grateful to him for his advice, which contributed significantly to the success of this work.

DOI: https://doi.org/10.5282/jums/v9i1pp1211-1240

(Atkinson et al., 2019; Muro & Whiton, 2018), which can be observed, for example, when looking at the spatial distribution of venture capital investments (Florida & King, 2016; Florida & Mellander, 2016), could lead to growing social tensions and the rise of political populism (Rodríguez-Pose, 2018). On the other hand, given the positive impact of entrepreneurial activity on economic performance (Isenberg, 2010; Szerb et al., 2015), dynamic and resilient EEs could be a means for smaller regions to close the gap with superstar regions in terms of job attractiveness, which could prevent phenomena such as brain drain, i.e., the outflow of human capital (see Alston, 2004). Since productive entrepreneurial activity in second-tier regions is likely to improve socioeconomic factors (Robinson et al., 2004), it may also have a positive impact on urban development efforts. Therefore, attempting to establish, grow, and sustain a local EE is a plausible strategy for improving the socioeconomic conditions of a second-tier region. Indeed, there are a growing number of thriving EEs in second-tier regions. Examples that have been analyzed in the literature include the U.S. cities of Chattanooga (Motoyama et al., 2016), Boise (Mayer, 2011), and Newton Falls (Roundy, 2019), and the Canadian cities of Waterloo and Calgary (Spigel, 2017).

Research on EEs often takes a static approach (Borissenko & Boschma, 2016). However, as EEs go through different stages of development (see Mack and Mayer, 2016), the effectiveness of specific support mechanisms depends on the state of an EE. Therefore, it is important to take a dynamic rather than a static approach to EE support and related public policy.

1.2. Research Question

This article addresses the impact of policies on the development of EEs in second-tier European regions. Thriving EEs can lead to improvements in socioeconomic factors, which is one reason why policymakers are increasingly interested in the concept (Roundy, 2017). At the same time, crafting and implementing policies that effectively contribute to a vibrant local EE can be challenging. Hence, the research question that forms the basis for the subsequent literature review and qualitative primary research is as follows:

What policies can second-tier European regions adopt, consistent with their stage of development, to effectively foster a local entrepreneurial ecosystem?

The research question consists of three main components that influence the scope of the research: the focus on Europe, the focus on second-tier regions, and the dynamics of policy measures in relation to the evolutionary development of an EE. The literature review in section 2 considers these three components in detail.

2. Literature Review

The second section provides a literature review of various aspects of an EE, including its foundations, relationship to economic and urban development, and the importance of spatial features. After developing the theoretical foundations of an EE, the literature review highlights the specific facets of the research question, i.e., the focus on Europe, the focus on second-tier regions, and the dynamics of an EE.

2.1. Foundations of Entrepreneurial Ecosystems

According to Roundy et al. (2018), Bahrami and Evans (1995) were the first to compare the Silicon Valley technology cluster to a natural ecosystem. Two years earlier, Moore (1993) stated more generally that firms are embedded in an ecosystem and do not develop in a vacuum. Early work introducing the term "entrepreneurial ecosystem" by Cohen (2006) and practical suggestions for developing an EE by Isenberg (2010) have contributed to the concept of EEs (Kuckertz, 2019) becoming known and attracting interest in both academia and practice (Roundy, 2017). While there are several definitions of the term "entrepreneurial ecosystem" in the literature (for an overview, see Cavallo et al., 2019), according to Cavallo et al. (2019, p. 1300), Stam's (2015) definition "[...] has been widely endorsed in literature [...]" and encompasses essential characteristics of an EE:

> "*The entrepreneurial ecosystem as a set of interdependent actors and factors coordinated in such a way that they enable productive entrepreneurship within a particular territory*" *(Stam, 2015, p. 1765)*.

A second definition that emphasizes dynamic processes within an EE by mentioning both the creation and growth of startups comes from (Spigel, 2017):

> "*[An entrepreneurial ecosystem is] a combination of social, political, economic, and cultural elements within a region that support the development and growth of innovative startups and encourage nascent entrepreneurs and other actors to take the risks of starting, funding, and otherwise assisting high-risk ventures*" *(p. 50)*.

Isenberg (2011a, 2011b) highlights that an EE consists of hundreds of specific elements that can be grouped into six larger areas: Culture, Policy, Finance, Human Capital, Support, and Markets. Isenberg also mentions that the combination of elements varies from EE to EE, indicating the uniqueness of each ecosystem, which in turn requires a specific rather than a generic approach to analyzing EEs. Furthermore, Isenberg (2011b) argues that the interaction of elements is highly complex and idiosyncratic, which limits the value of identifying generic causal relationships. An illustration of Isenberg's understanding of an EE can be found in Figure 1, and a more detailed elaboration of the domains can be found in Appendix 1.

According to Mack and Mayer (2016), it is necessary to consider the evolutionary dynamics of the components of an EE as well as its developmental stages. Several other authors share this view (e.g., Isenberg, 2011b; Kuckertz, 2019; Mason and Brown, 2014). Mack and Mayer (2016) divided

the development of an EE into four stages: Birth, Growth, Sustainment, and Decline. Each of these phases has differences in the number of firm entries and firm exits. Mack and Mayer's development model can be seen in Figure 2. The full figure with some of the effects of the phases can be found in Appendix 2.

Based on the two definitions presented in combination with Isenberg's (2011b) understanding of EE domains and Mack and Mayer's (2016) evolutionary development phases, EEs share some common characteristics, including:

- the interdependence of actors and factors,
- the need for coordination and processes to enable productive entrepreneurship,
- the focus on a particular territory, and
- the need for an evolutionary and dynamic perspective.

After discussing the definition and characteristics of an EE, its major areas and phases of development, section 2.2 focuses on the impact of an EE on the economic development of the area in which it is located.

2.2. Entrepreneurial Ecosystems and Economic Development

A positive relationship between entrepreneurial activity and economic prosperity has been noted by several scholars (e.g., Audretsch et al., 2015; Feldman, 2014; Isenberg, 2011b). In contrast to some authors who focus predominantly on the impact of high-growth startups (e.g., Mason and Brown, 2014), Stam (2015) argues that this focus is too exclusive, as other innovative startups and entrepreneurial workers can also contribute to positive welfare outcomes to some extent. Stam refers to an article by Baumol (1990) that distinguishes between productive, unproductive, and destructive entrepreneurial activities and their respective effects on an economy's productivity growth. Several factors influence the extent to which entrepreneurship contributes to the economy of a given area, with some scholars emphasizing that not all entrepreneurial activities contribute positively to its development (e.g., Mason and Brown, 2014; Nightingale and Coad, 2014). Rather, there are a small number of high-growth startups that contribute positively to overall economic growth and many underperforming firms that "[...] have low productivity and low levels of innovation, and generate churn rather than economic growth" (Nightingale & Coad, 2014, p. 130). In addition, N. Lee and Rodríguez-Pose (2021) found that entrepreneurship can lead to regional poverty reduction when it occurs in tradable sectors, increasing the likelihood of positive regional multiplier effects. Entrepreneurship in non-tradable sectors still has some economic benefits despite the likelihood of local market saturation, but is not sufficient to contribute significantly to regional poverty reduction (N. Lee & Rodríguez-Pose, 2021). Aparicio et al. (2020) emphasize that entrepreneurship, if indeed productive, can be a vehicle for inclusive growth that includes a region's vulnerable communities.

Policymakers can achieve better outcomes by distinguishing between different forms of entrepreneurial activity and taking a systems-based, holistic approach rather than focusing on firm-specific interventions (Mason & Brown, 2014). EEs represent one such systemsbased approach. Several scholars have linked EEs to economic growth (e.g., Isenberg, 2011b; Mason and Brown, 2014). This is driven by the emergence of high-growth startups that are enabled by a supportive ecosystem and contribute disproportionately to the economic development of a region (Mason & Brown, 2014; Nightingale & Coad, 2014), e.g., by creating a significant share of new jobs compared to non-high-growth firms (Coad et al., 2014; Henrekson & Johansson, 2010) and by contributing to the growth of other firms in the same area through knowledge spillovers (Acs et al., 2009; Mason & Brown, 2014). However, this growth takes time, so policymakers need to take a long-term perspective (Nightingale & Coad, 2014).

The arguments presented in this subsection seem to motivate policymakers to promote entrepreneurial activities and establish local EEs. However, an EE may also have implications for urban development efforts in each area, touching on a domain that is generally not directly related to entrepreneurship. The following subsection analyses this relationship.

2.3. Entrepreneurial Ecosystems and Urban Development

Some scholars argue that Jane Jacobs, one of the world's most influential urbanists, was quick to emphasize that cities' ability to attract diverse people provides fertile ground for creativity and innovation, which are key to entrepreneurial activity (Hospers, 2006; S. Y. Lee et al., 2004), economic growth (Florida, 2003, p. 43), and vitality (Auerswald, 2015). Because of the density of labor, capital, knowledge, and other resources, cities enable the creation of young firms, a concept Jacobs calls "new work" (Jacobs, 1969, p. 49ff.). Her concept of new work in the context of urban development therefore created some of the frameworks for EEs found in the modern literature. Moreover, she also described the importance of knowledge spillovers between different industries for the economic growth of cities (Desrochers & Hospers, 2007; Qian, 2018). Indeed, Florida et al. (2017, p. 92) state that "[c]ities [...] are the enabling infrastructure where connections take place, networks are built[,] and innovative communications are consummated." In line with Jacobs' observations, Richard Florida's theory of the creative class suggests that creativity and an open culture are "[...] a spur to societal innovation, entrepreneurship, and economic development" (Florida, 2005a, p. 6). Combining Jacobs' argument of the attractiveness of cities to a variety of people and the availability of resources in cities with Florida's theory of the creative class illustrates the link between EEs and urban development. Florida et al. (2017) summarize this finding by stating that innovation, entrepreneurship, and creativity are essentially spatial rather than individual or firm-level processes.

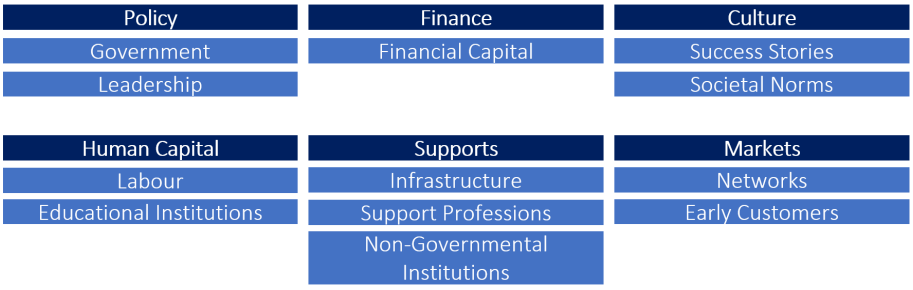


Figure 1: Domains of an Entrepreneurial Ecosystem. Source: adapted from Isenberg (2011b).

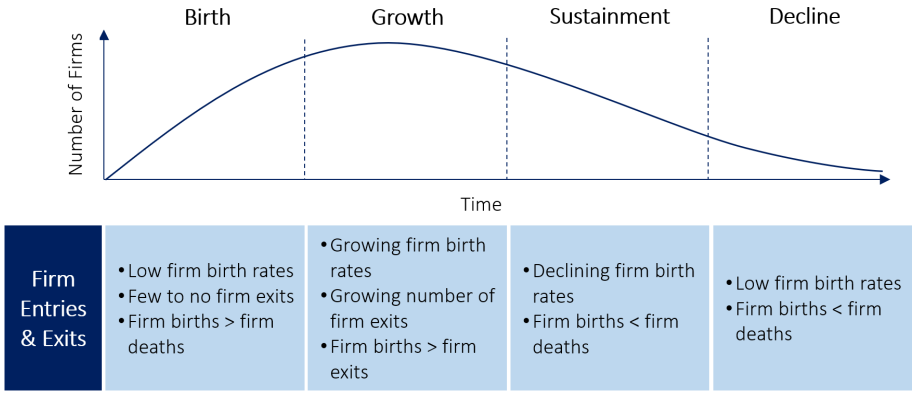


Figure 2: Evolution of an Entrepreneurial Ecosystem. Source: adapted from Mack and Mayer (2016).

As discussed in section 2.2, establishing a thriving EE can contribute to local economic development. This, in turn, has the potential to both increase tax revenues and improve the availability of private capital in a region. As Jacobs (1969, pp. 290-317) noted, urban development in each area requires financial resources, which can come from public or private sources. Therefore, the establishment of a local EE could be part of a broader urban development strategy if an increase in public and/or private funding is anticipated. Similarly, Welter et al. (2008) suggest that promoting entrepreneurship in distressed urban areas, if certain barriers can be overcome, may contribute to their regeneration. As Isenberg (2011b) has noted, each EE is embedded in the context of a particular area and is therefore made up of a unique combination of different elements that interact in complex and idiosyncratic ways. Thus, when attempting to create a local EE, spatial characteristics must be considered.

The section 2.4 highlights the differences between these features and analyzes some of their implications. This sets the stage for the main concern of this article: the unique context and configuration of EEs in second-tier European regions, and the corresponding implications for policy effectiveness throughout the life cycle of an EE.

2.4. Entrepreneurial Ecosystems and Spatial Differences

In this subsection, spatial differences are introduced into the discussion of EEs. First, a phenomenon that affects the geographic inequality of EEs is explained: the polarization of innovation. Then, the focus shifts from a more general consideration of EEs to EEs specifically in second-tier regions. A brief analysis of the uniqueness of the contemporary European context is also provided. This subsection thus introduces the focus of the article: the development of EEs in second-tier European regions.

2.4.1. Polarization of Innovation

In 2005, Richard Florida pointed out in an article in *The Atlantic Monthly* that contrary to the widespread belief that technology and globalization are leveling the global economic playing field, the world has only a few centers where innovation is concentrated, and that this divide is also evident at the national level (Florida, 2005b). He notes that "[p]opulation and economic activity are both spiky, but it's innovation - the engine of economic growth - that most concentrated" (p. 49). Analysis of venture capital investment (VC) in U.S. metropolitan regions, European metropolitan regions, and metropolitan regions around the world (Florida & King, 2016) confirms Florida's original observation of innovation concentration. Florida and King (2016, pp. 6-7) note that only ten metropolitan regions worldwide account for more than half of total global VC investment. Areas in the U.S. accumulated 68.6% of global VC investment, with Silicon Valley alone attracting 25.3% (Florida & King, 2016, pp. 6-11). In the following, this phenomenon is referred to as the polarization of innovation. Figure 3 illustrates the distribution of VC investments worldwide. Appendix 3 illustrates the distribution within the U.S. and Europe in detail.

According to Atkinson et al. (2019), the polarization of innovation leads to a growing gap between superstar metropolitan areas and non-superstar areas, as well as a

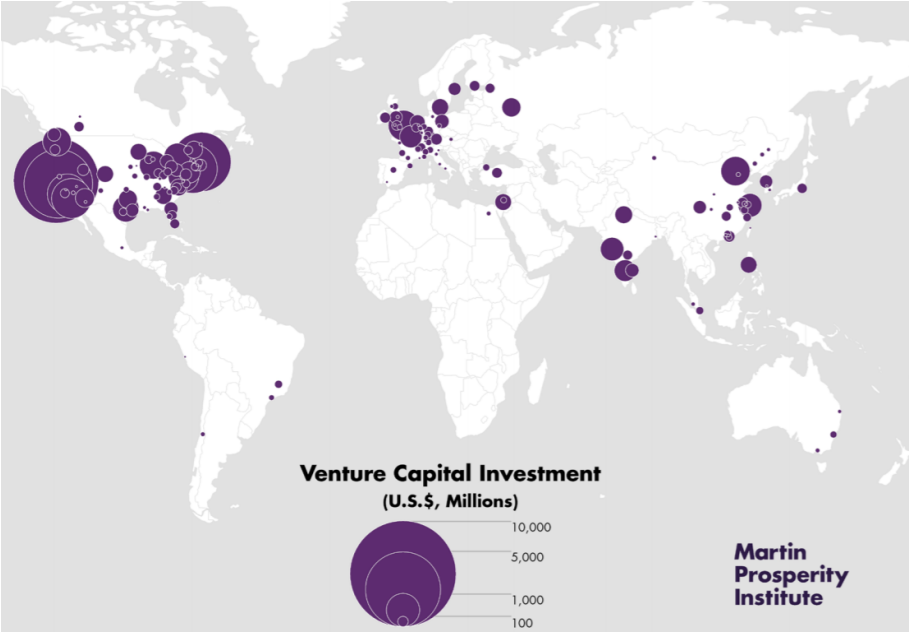


Figure 3: Venture Capital Investments across Metropolitan Areas Worldwide in 2012. Source: Florida and King (2016, p. 10).

concentration of jobs in the innovation sector. This leads to, among other things, higher housing prices and more traffic in superstar areas and out-migration, territorial underdevelopment, and economic exclusion in second-tier and rural areas (Atkinson et al., 2019). Rodríguez-Pose (2018) suggests that such regional disparities even lead to social tensions and political populism based on territorial rather than social foundations.

As discussed in section 2.2, EEs can contribute to the economic development of an area. While most thriving EEs are located in superstar areas, some second-tier regions around the world have managed to build thriving EEs despite innovation polarization. section 2.4.2 focuses on EEs in second-tier regions, while section 2.4.3 highlights the uniqueness of the contemporary European context, which has not yet received much attention in the entrepreneurship literature (Audretsch et al., 2015).

2.4.2. Entrepreneurial Ecosystems in Second-Tier Regions

There are numerous definitions of second-tier cities or regions in both academic and practitioner literature (see Figure 4 for a non-exhaustive overview). In this article, second-tier regions are defined as regions "[...] that are smaller than the large metropolises that dominate regional or national economies [...]" (Wachsmuth, 2008, p. 1), although no minimum or maximum population size is specified, as second-tier regions should "[...] be defined in relation to the first-tier cities with which they coexist" (Wachsmuth, 2008, p. 2).

As Isenberg (2011b) has noted, the design of an EE depends on the local context and conditions. Therefore, it is neither advisable nor feasible to emulate superstar EEs like Silicon Valley, especially for second-tier regions. Instead, it should be recognized that each EE is different and has unique characteristics (Xu & Dobson, 2019). Peripheral places face certain challenges related to their spatial context and availability of resources as opposed to superstar areas (Xu & Dobson, 2019). These challenges may include more difficult access to financial resources and skilled human capital, a less supportive sociocultural environment, lack of certain infrastructure components (such as transportation infrastructure or high-speed internet), limited markets and market access, and ineffective policies (Xu & Dobson, 2019).

In contrast to the innovation polarization phenomenon, and despite the above challenges, there are some examples in the literature of second-tier regions that have successfully built thriving EEs. Such examples include Boise, Portland, and Kansas City (Mayer, 2011), Chattanooga (Motoyama et al., 2016), Newton Falls and Geneva (Roundy, 2019), and Calgary and Waterloo (Spigel, 2017). Summaries of the above case studies can be found in Appendix 4.

2.4.3. European Context

Compared to North America, the contemporary urban context in Europe has been less studied in the academic literature (Audretsch et al., 2015). However, Europe has many small and medium-sized cities, highlighting its unique polycentric urban structure (Dijkstra et al., 2013). Dijkstra et al. (2013) argue that stylized analytical frameworks, assumptions, and policy conclusions derived from them are mostly based on the contexts of North America or the developing world and therefore have limited relevance to the contemporary European context. This in turn hinders the process of policy development in European regions based on textbook models (Dijkstra et al., 2013). Given the importance of spatial context in discussing EEs and the focus of the academic literature on North America and developing countries, this

Source	Definition
Markusen & DiGiovanna (1999, p. 3)	"[...] spatially distinct areas of economic activity where a specialized set of trade-oriented industries takes root and flourishes, establishing employment and population growth trajectories that are the envy of many other places."
Wachsmuth (2008, pp. 1-2)	"The terms 'second-tier city' and 'third-tier city' are frequently used to describe urbanized areas that are smaller than the large metropolises that dominate regional or national economies, but they have no single or accepted definition. In fact, this fluidity is inherent in the concept [...] and [...] can only be defined in relation to the first-tier cities with which they coexist."
Mayer (2011, p. 7)	"Second tier high-tech regions [...] specialize in certain sub-sectors of the high-tech economy and take advantage of the presence of lead high-tech firms."
Parkinson et al. (2012, p. 3)	"[...] cities outside the capital whose economic and social performance is sufficiently important to affect the potential performance of the national economy."
Dijkstra et al. (2013, p. 354)	"Second-tier metro regions are the group of largest cities in the country excluding the capital."
Cardoso (2018, p. 223)	"[...] non-capital, medium-sized cities, which do not play dominant roles in their countries but are economically, culturally and demographically significant [...]."

Figure 4: Selection of Definitions of Second-Tier Regions.

article considers the unique European context as well as the individual contexts of regions within Europe.

Audretsch et al. (2002, p. 4) argue that the U.S. internalized the virtues of entrepreneurship more quickly than Europe and that European countries were relatively slow to adopt a similar mindset. European attitudes toward the entrepreneurial economy developed in five phases (for a summary, see Audretsch et al., 2002, pp. 4-6). Toward the end of the 1990s, European policymakers reached a consensus on the superiority of the entrepreneurial economy in the United States over the old managerial economy in Europe, leading to a commitment to create a new European entrepreneurial economy (see European Commission, 2000, pp. 249-286). In 2013, in response to the 2008 financial crisis, the European Commission presented an action plan to reignite entrepreneurship in Europe through governance mechanisms (European Commission, 2013, p. 3). Szerb et al. (2020) explain that since then, a policy priority setting framework, the Research and Innovation Strategies for Smart Specialization (RIS3) agenda, has emerged in the EU. The RIS3 agenda aims to tailor R&D and innovation-related policies to the capabilities, strengths, and potential of a given region. According to some scholars, RIS3 is an innovative policy approach (Foray & Goenaga, 2013) that is part of a broader regional and place-specific growth policy framework (OECD, 2013; Pugh, 2014). Morgan (2017, p. 569) even describes it as "[...] the most ambitious innovation program ever introduced in the European Union [...]". Szerb et al. (2020) emphasize that the RIS3 agenda recognizes the spatial differences between regions within the EU, resulting in individual contexts that require tailored governance and policy approaches. The same authors then compare RIS3 with the EE approach. While both approaches respect spatial differences, EEs consider a broader range of individual and institutional factors and the interconnectedness of actors within an ecosystem (Acs et al., 2016; Szerb et al., 2020). Nevertheless, Szerb et al.'s (2020) analysis shows that spatial differences are increasingly considered by policymakers in Europe, both at supranational and national levels. Regarding the development of EEs and to fully understand the framework conditions of each region, identify institutional and individual weaknesses, understand the harmonization of the components of an EE, and simulate policies that could alleviate bottlenecks of the regional EE, Szerb et al. (2020) argue that the use of the Regional Entrepreneurship and Innovation Index (REDI) is appropriate (Szerb et al., 2020; see also Figures 5 and 6). For more information on the REDI, see section 2.5, which also highlights the identified research gap.

2.5. Research Gap

Since the early work of scholars such as Cohen (2006), Isenberg (2010), and Feld (2020), many aspects of EEs have been explored in the academic literature. However, research on EEs has mostly focused on nations (Audretsch et al., 2015) or on established ecosystems in large superstar regions (Roundy, 2019). As a result, there are still a wealth of areas that can be further explored in academic research (Cavallo et al., 2019). This article aims to outline policy implications for second-tier European regions that aim to build, grow, and sustain a resilient EE. The importance of a dynamic

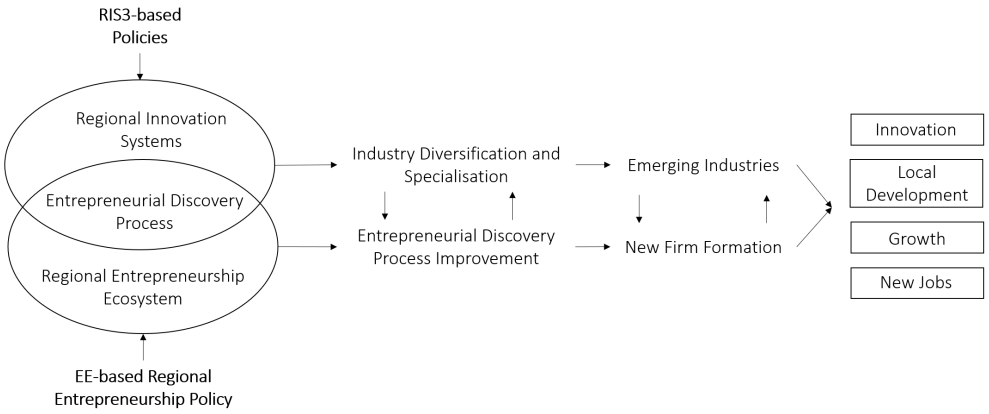


Figure 5: Joint Potential of RIS3- and EE-Based Regional Entrepreneurship Policy. Source: adapted from Szerb et al. (2020, p. 7).

Structure of the REDI: 3 sub-indices/14 pillars		National and regional institution variables	Regional level individual variables
Entrepreneurial Aspiration Sub-index	Financing	Financial Institutions	Informal Investment
	Globalisation	Connectivity	Export
	High Growth	Clustering	Gazelle
	Process Innovation	Technology Development	New Technology
	Product Innovation	Technology Transfer	New Product
Entrepreneurial Ability Sub-index	Competition	Business Strategy	Competitors
	Human Capital	Education and Training	Education Level
	Technology Sector	Absorptive Capacity	Technology Level
	Opportunity Start-up	Business Environment	Opportunity Motivation
Entrepreneurial Attitudes Sub-index	Cultural Support	Open Society	Carrier Status
	Networking	Social Capital	Know Entrepreneurs
	Risk Acceptance	Business Risk	Business Acceptance
	Start-up Skills	Quality of Education	Skill Perception
	Opportunity Perception	Market Agglomeration	Opportunity Recognition

Figure 6: Structure of the Sub-Indices and Pillars of the REDI. Source: adapted from Szerb et al. (2020, p. 9).

approach to research EEs and its underdevelopment in the scientific literature has been noted by several researchers. Both Borissenko and Boschma (2016) and Mack and Mayer (2016) emphasize the need for further research on the dynamics of EEs. Cavallo et al. (2019) emphasize the need to explore the role of policymakers in the dynamic evolution of an EE and how to promote "natural" and self-regulating rather than artificial mechanisms. A call for papers for *Entrepreneurship & Regional Development* by Audretsch et al. (2018) identified the need for papers that address policy issues, such as analyzing how policy influences elements of an EE and how elements of an EE influence policy. Shwetzer et al. (2019, p. 89) additionally articulated two research directions that align with the above pathways for further research, namely "[p]olicymakers' interventions and support to enable and grow EEs" and "EEs creation, growth and how can they be sustained". By elaborating policy implications for European second-tier regions while considering the dynamics of an EE, this article aims to contribute to filling the research gaps mentioned before. By focusing specifically on second-tier European regions, this article also considers the uniqueness of the European context (Audretsch et al., 2015).

After an introduction to the context and challenges European second-tier European regions in establishing an EE, a comparative case study is provided of two second-tier European regions that, contrary to expectations, have managed to establish and develop thriving EEs. The selection of cases is based on the REDI, as it is comparable for all EU regions (Szerb et al., 2013; Szerb et al., 2015). The REDI was constructed to "[...] capture the contextual features of entrepreneurship across EU regions" (Szerb et al., 2015, p. 1) and combines three sub-indices, 14 pillars, and both individual and institutional variables (see Figure 6). Based on these criteria, Szerb et al. (2015, p. 14) created a ranking of regional entrepreneurship that includes the 125 regions of the European Union. The ranking of the top 25 regions can be seen in Figure 7, the original full ranking table from Appendix 5.

Using the REDI, it is possible to identify the second-tier regions that have succeeded in establishing competitive EEs. The regions ranked first through fourth are all superstar regions. However, there are also some regions in the top ten that are not major cities (defined as regions with more than 250,000 inhabitants, according to Audretsch et al., 2015), including:

- 5th place: East Middle Sweden (largest city: Uppsala)

Rank	Code	Region	REDI
1	DK01	Hovedstaden	82.2
2	UKI	London	79.9
3	FR1	Île de France	79.2
4	SE11	Stockholm	73.8
5	SE12	Östra Mellansverige	72.7
6	SE23	Vastsverige	72.2
7-8	IE02	Southern and Eastern	72.0
7-8	DK05	Nordjylland	72.0
9	UKJ	South East (UK)	69.5
10	SE22	Sydsverige	67.3
11	DE03	Berlin	67.2
12	DK03	Syddanmark	65.1
13	BE1	Région de Bruxelles-Capitale	64.9
14	SE33	Övre Norrland	64.7
15	NL3	West-Nederland	64.4
16	DK04	Midtjylland	64.3
17	FR7	Centre-Est (FR)	64.2
18	IE01	Border, Midland and Western	63.4
19	DE7	Hessen	63.3
20	FI1B	Helsinki-Uusimaa	62.2
21	BE2	Vlaams Gewest	62.1
22	UKH	East of England	61.5
23-25	DK02	Sjalland	60.7
23-25	UKK	South West (UK)	60.7
23-25	AT1	Ostösterreich	60.7

Figure 7: REDI Ranking and REDI Scores of the Top 25 European Union Regions. Source: adapted from Szerb et al. (2015, p. 14).

- 7th–8th place: Denmark's North Jutland Region (largest city: Aalborg)
- 9th place: Great Britain's South East (largest city: Brighton and Hove)

When looking at cities with less than 100,000 inhabitants, i.e., small cities (in line with Audretsch et al., 2015), the following regions ranked in the first quarter also seem promising for a more detailed analysis:

- 14th place: Sweden's Upper Norrland (largest city: Umeå)
- 18th place: Ireland's Border, Midland and Western (largest city: Galway)
- 23rd–25th place: Denmark's Zealand Region (largest city: Roskilde)

The author of this article decided to examine the cases of East Middle Sweden and Border, Midland and Western (Ireland) to identify specific factors that contribute to the regions' entrepreneurial success. East Middle Sweden is an interesting case because the region is near Stockholm, the capital of Sweden, yet has been able to develop its own EE, which has produced prominent companies such as Skype and Klarna (see section 4.1 for the detailed analysis). The case of Border, Midland and Western is interesting because it is in a rather rural region on the west coast of Ireland, with an underdeveloped transport infrastructure hampering domestic and international travel (see section 4.2 for a detailed analysis). For linguistic and readability purposes, East Middle Sweden will be referred to as

"Uppsala" and Border, Midland, and Western as "Galway" in the following; however, the surrounding areas as integral parts of the respective EE remain included in the analysis.

After defining the research gap and the two regional EEs that will be analyzed in depth, section 3 describes the research methodology used to analyze the development of the EEs in Uppsala and Galway.

3. Research Method

This section describes the research method chosen for this article, including the research approach, the process of data collection, and the analysis of the data. A more detailed explanation can be found in Appendix 6.

3.1. Research Approach

An inductive research approach was adopted to analyze the dynamic evolution of EEs in second-tier European regions. This approach consists of guiding research questions that determine data collection and analysis (Roundy, 2019). Because there is little research on EEs in second-tier European regions (Roundy, 2019), an inductive, theory-building approach was appropriate (Locke, 2007). Inductive theory building allows the researcher to gain an understanding of the unfolding of processes behind unusual phenomena and explain surprising occurrences (Edmondson & McManus, 2007). The emergence and growth of an EE in a second-tier European region represent such complex processes with multiple variables to consider, such as time, social interactions, and feedback loops (Roundy, 2019).

By choosing a comparative case study approach, similarities and differences in the context, developmental dynamics, and composition of the two EEs in Uppsala and Galway could be derived. While such a "small n" approach (Roundy, 2019) limits generalizability of findings, some scholars argue that comparative case studies generate some degree of generalizability despite a small sample size (e.g., Eisenhardt and Graebner, 2007; Roundy, 2019). However, given the exploratory nature of the research, the goal of this article is particularization rather than generalization (Welch, 2011).

3.2. Data Collection

To gain a deeper understanding of the processes underlying the development of the EEs in Uppsala and Galway, interviews were conducted with representatives from the two

regions. Purposive sampling, a form of non-probability sampling, was chosen to select the interviewees. Given the research objectives and inductive research approach, the sampling procedure did not aim to be statistically significant, but rather to select information-rich cases to gain insight into the nature, processes, and structure of the two EEs (Patton, 2005). Interviewees were selected to allow for the inclusion of different perspectives, knowledge, and experiences of the various stakeholders that comprise an EE (see Isenberg, 2011b; Palinkas et al., 2015). The respective interviewees were identified through a combination of secondary data highlighting key actors within each EE and snowballing (see, e.g., Biernacki and Waldorf, 1981).

The predominant determinant of sample size in qualitative research is information saturation (e.g., Fusch and Ness, 2015; Malterud et al., 2016; Morse, 1995). Because some determinants of information saturation could not be defined a priori, it was not possible to determine a final number of interviews to be conducted at the beginning of the research process. However, as Malterud et al. (2016) mention, a first approximation of the sample size is necessary for research design. Therefore, the author's goal was to conduct at least five interviews in each of the two regional EEs to include the knowledge, views, and experiences of key stakeholders in the EEs. However, the final sample size had to be continuously evaluated throughout the research process and was based on the informativeness of the interviews conducted in relation to the research objectives (Malterud et al., 2016). The final sample table can be found in Appendix 6. The interviews conducted were semi-structured. While a semi-structured interview is more conversational, a list of predetermined questions was used, which provided the opportunity to explore topics in greater depth if deemed important for later analysis of each EE (Longhurst, 2003). The guide used for the semi-structured interviews can be found in Appendix 9.

3.3. Data Analysis

The research for this article was divided into three phases (see Figure 8). Interviews as part of the second phase were conducted via video or phone call due to travel restrictions resulting from the Covid-19 pandemic. To facilitate effective data analysis, verbal transcripts (Matheson, 2007; McLellan et al., 2003; Wellard & McKenna, 2001) were prepared using standard orthography (Kowal & O'Connell, 2004, pp. 249-250). The transcripts can be found in Appendix 11.

The process of coding the interviews followed the suggestions of Schmidt (2004, pp. 254-257). First, the common themes discussed during the interviews were identified and summarized in an analytical guide (see Appendix 10). Then, the interviews were coded accordingly. Based on the coded interviews and by grouping the quotes according to their codes, the relevance of each theme was assessed. Depending on the relevance of the themes, each interview was iteratively analyzed again in depth to understand the meaning of the different interviews and to reach a conclusion as to why each of the two regions studied achieved the establishment of a thriving EE. In addition, following the example of some researchers (e.g., Motoyama et al., 2016; P. Ryan et al., 2021), timelines were used to analyze relevant innovation pathways and the dynamics of each EE. By sequencing secondary data points and adding a temporal layer to the interview coding process, both the creation and subsequent analysis and interpretation of timelines and relevant innovation trajectories became possible.

Interpretive sensemaking was used to analyze the data on Uppsala and Galway. Interpretive sensemaking avoids generalizations and emphasizes the importance of individual context (Welch, 2011). By sequencing relevant innovation trajectories, process-based interpretation, and explanation of the evolutionary dynamics of the two EEs became possible (P. Ryan et al., 2021). The analysis was refined through successive iterations between theory and data (G. W. Ryan & Bernard, 2000, p. 783). By combining the viewpoints and perspectives of the various interviewees, as well as data obtained from secondary sources, "chain[s] of evidence and narrative accounts" (P. Ryan et al., 2021, p. 7) emerged that reveal the evolutionary processes of the two EEs in Uppsala and Galway. The interpretation of the data always considered the spatial context of each case (Roundy, 2019). Figure 9 summarizes the data sources used and their analytical use for this article. The next section presents the results.

4. Findings

This section summarizes the results of the primary and secondary data collected on the EEs in Uppsala and Galway. It also summarizes the similarities and differences between the two EEs. The underlying data structures are presented towards the end of each stage of development; the exemplary quotes can be found in Appendix 8.

4.1. Analysis of Uppsala's Entrepreneurial Ecosystem

Due to its extensive academic history dating back to 1477, Uppsala is primarily known as a university town, and research on Uppsala's economy has focused primarily on the life sciences cluster (e.g., Teigland et al., 2004, 2007; Waxell and Malmberg, 2007). The local EE has not received much attention in scientific research. Therefore, information about Uppsala's EE comes primarily from qualitative interviews with stakeholders within the EE unless otherwise noted. The context of Uppsala is summarized in Figure 10.

4.1.1. Birth: Research Breakthroughs at Uppsala University

The roots of entrepreneurial activity in Uppsala can be traced to significant research findings in biotechnology at Uppsala University in the 1920s and 1930s (Waxell & Malmberg, 2007). These findings led to a growing interest in university-industry collaboration in the life sciences. In 1950, the large Swedish pharmaceutical company Pharmacia decided to move its core business from Stockholm to Uppsala (Eliasson & Eliasson, 2006; Waxell & Malmberg, 2007). The reason for this was the successful R&D collaboration with Uppsala University (Eliasson & Eliasson, 2006), which mainly

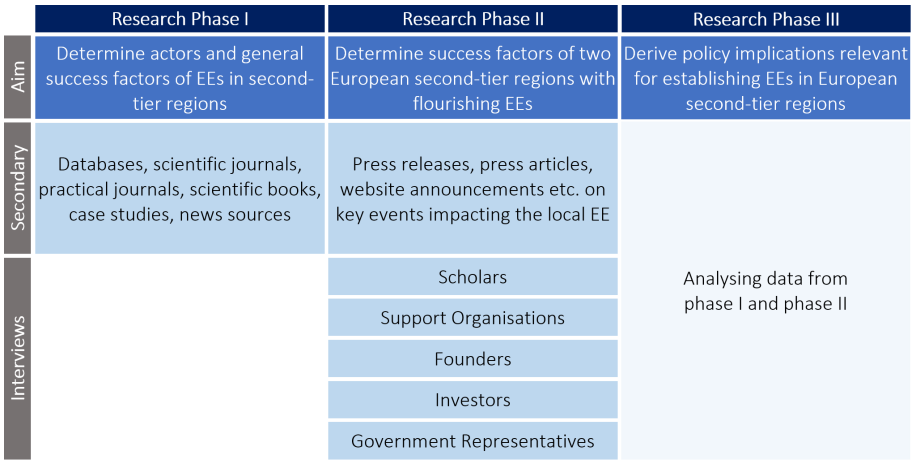


Figure 8: Research Phases. Source: own illustration.

Data Source	Type of Data	Analytical Use
Primary Sources		
Interviews	**Semi-structured interviews in Uppsala (9):** • 1 local government representative • 1 scholar from a local university • 2 representatives from a major support organisation • 1 representative from an industry-specific support organisation • 2 entrepreneurs located in Uppsala • 1 entrepreneur located in Stockholm with activities in Uppsala • 1 representative from an investor network **Semi-structured interviews in Galway (5):** • 1 local government representative • 1 scholar from a local university • 1 representative involved in two major support organisations and a co-working facility • 1 investor and simultaneous representative of a major support organisation, a co-working facility, and a local business network • 1 entrepreneur located in Galway	• Reconstructing the historical development of the EE • Evaluating economic trajectories and key events with an impact on the development of the EE • Mapping the current state of the EE and its context • Evaluating key actors, assessing their roles within, and their importance for, the EE • Outlining the strengths and weaknesses, as well as their (contextual) root causes, of the EE
Secondary Sources		
Scientific papers	Multiple scientific papers on key economic development trajectories and initiatives in Uppsala and Galway	• Reconstructing the historical development of the EE • Understanding the underlying processes of the respective EE's development in Uppsala and Galway • Providing support to data and interpretation
Websites of regional organisations and press coverage	Websites of relevant actors within the respective regional EE and multiple press clippings regarding the development and the state of the regional EE	• Understanding the perception of the respective region in a local, national, and international context • Providing support to the interpretation of economic trajectories and key events • Providing support to the role assessment of key actors

Figure 9: Data Sources and Analytical Use. Source: own illustration.

revolved around two Nobel Prize winners in chemistry (Waxell & Malmberg, 2007). Pharmacia then gradually became one of the most important employers in the region and contributed to the growth of the Uppsala life science cluster.

After Pharmacia moved its headquarters to Uppsala, the company continuously engaged in significant M&A activity (Dahlgren & Valentin, 2007). Then, in the mid-1990s, Pharmacia began spinning off, selling, and restructuring various parts of the company (Waxell & Malmberg, 2007), which attracted two large multinationals to the region (Eliasson & Eliasson, 2006): Pfizer and Cytiva (formerly GE Healthcare). Uppsala's attractiveness to life science companies can also be attributed in part to the increased availability of scientifically trained and industrially experienced people following the dissolution of Pharmacia (Eliasson & Eliasson, 2006), supporting Mayer's (2011, p. 205) observation that MNEs act as "surrogate universities".

The presence of Pharmacia, and later other multina-

Context of Uppsala's Entrepreneurial Ecosystem	
Location	East Middle Sweden (~70 km north of Stockholm, ~40 km north of Arlanda Airport)
Population	City of Uppsala: ~172,000 | Uppsala County: ~367,000
EE Size	378 start-ups in Uppsala County listed on Crunchbase[1]
Funding	USD 361.9m (104 funding rounds) since 2000 listed on Crunchbase[1]
Industry Cluster	Biotechnology (Life Science)
Success Stories[2]	Skype (ICT) | Klarna (FinTech) | MySQL (ICT) | Orexo (Pharmaceuticals)
Universities	Uppsala University (UU) | Swedish University of Agricultural Sciences (SLU)
Support Organisations[2]	STUNS | UU Innovation | Drivhuset
Incubators / Accelerators[2]	Uppsala Innovation Centre
Coworking Spaces[2]	Uppsala Innovation Hub | Green Innovation Park | BASE10
Financial Capital[2]	Almi | Connect Uppsala | UU Holding | SLU Holding | EASME
Notable MNEs[2]	Cytiva (former GE Healthcare) | Fresenius Kabi | Thermo Fisher Scientific

[1] Crunchbase is increasingly being used as a database for scientific research in economics and management (Dalle et al., 2017), which is why its data is used to provide an overview of Uppsala's EE. Access date: 20 July 2020.
[2] Main actors according to interviewees and secondary data; therefore the list is non-exhaustive.

Figure 10: Case Context of Uppsala. Source: own illustration.

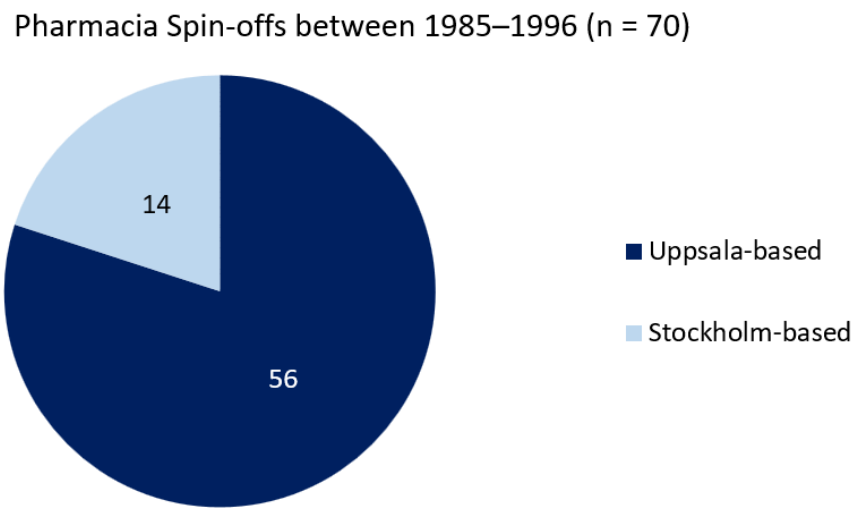


Figure 11: Pharmacia Spin-offs between 1985-1996. Source: own illustration based on Eliasson and Eliasson (2006) and data from Nilsson and Norell (1996).

tionals, in Uppsala has encouraged the creation of spin-offs by former employees of these firms ((Eliasson & Eliasson, 2006); see also Appendix 7 for a non-exhaustive overview of Pharmacia's Swedish genealogy). For example, Pharmacia established 70 companies between 1985 and 1996 (see Figure 11), 56 of which were located in the Uppsala region and 14 in the Stockholm region (Eliasson & Eliasson, 2006), again underscoring Uppsala's attractiveness to the life sciences industry. Even when Pfizer, which acquired parts of Pharmacia during the breakup period, later phased out the research activities started by Pharmacia in Uppsala, former Pharmacia employees continued to establish new companies in the region that were not based on technologies developed at Pharmacia (Eliasson & Eliasson, 2006). For a detailed recapitulation of Pharmacia's history and its aftermath, see Eliasson and Eliasson (2006).

However, the entrepreneurs in Uppsala did not only emerge from the local multinationals during the birth of the EE. The importance of the academic sector in the city ensured a steady influx of domestic and foreign students, as well as professors and researchers. These groups of people were of great importance for the emergence of Uppsala's EE. It is noteworthy that the Swedish legal framework provides university researchers with rights to their results, which is known as "professor's privilege" (see, e.g., Färnstrand Damsgaard and Thursby, 2013). Thus, unlike other countries, researchers at Swedish universities are allowed to commercialize their results without having to deal with legal issues related to intellectual property (IP). According to several interviewees, this mechanism, combined with an emerging support system focused primarily on the life sciences cluster, encouraged professors and researchers to pursue entrepreneurial ventures. Interestingly, most professors and researchers remained employed part-time at the university, resulting in hybrid entrepreneurship (Folta et al., 2010).

4.1.2. Growth: Emergence of an Entrepreneurial Support System

Interviewees attributed the growth of Uppsala's EE to many aspects revolving around activities and initiatives either initiated or supported by government officials. Due to high tax rates and the general importance of the public sector in Sweden, several interviewees stated that the activities of entrepreneurial support organizations in Uppsala have been and continue to be at least partially dependent on public funding. Among government officials, successful university-industry collaboration in research and development has gradually awakened interest in the shared potential of academia, business, and the public sector, as well as in the potential of entrepreneurial activities. One of the major initiatives aimed at harnessing this potential was the creation of STUNS, an independent nonprofit organization dedicated to orchestrating the now emerging local EE. STUNS's founding in 1983 was based on a plan by the governor at the time to create a local science park modeled on Silicon Valley.

Third-level funding and the initiation of STUNS are just two examples of the emerging public interest in local entrepreneurship. As Uppsala grew, other public and semi-public organizations and actors gradually emerged. Examples include Almi, a regional development office that provides financial capital to companies; Connect, an organization that connects entrepreneurs with domestic and foreign investors; Drivhuset, a support organization for student entrepreneurs; and the Uppsala Innovation Centre, a leading international business incubator.

Although many different organizations are involved in Uppsala's EE, there has been no sign of competition between actors within the ecosystem during the growth phase, which has been and continues to be an important regional success factor. Close collaboration between actors within the ecosystem is a major strength of Uppsala, according to several interviewees. Because the city of Uppsala is spatially compact, there are certain geographic locations where entrepreneurial activities are concentrated. The resulting density of the EE facilitates knowledge exchange between different actors within the ecosystem and catalyzes entrepreneurial activity in the region. As one interviewee from STUNS noted:

> *"[...] since [...] Uppsala is a fairly small city, we try to have [...] a no-wrong-door policy. [...] we are a lot of actors within the innovation ecosystem that try to refer people depending on how mature their idea is [...]."*

Due to the above factors and processes that influenced the growth of Uppsala's EE, the region gained some notoriety as the birthplace of companies such as Skype, Klarna, MySQL and Orexo. Such success stories, several interviewees said, were fundamental to Uppsala's EE, e.g., by inspiring and motivating students to become entrepreneurs. Entrepreneurial role models are certainly crucial to the ecosystem even today.

The growth of Uppsala's EE is partly due to its spatial location. The city is close to Sweden's largest international airport, Arlanda, which has been particularly important to the local life science cluster given its global reach. The startups that emerged in the growth phase of the Uppsala industrial cluster, after serving regional markets in the initial phase of the EE, were therefore usually "born globals", i.e., they served several international markets early in their development (Knight & Cavusgil, 1996). In addition, public transportation between Uppsala and Stockholm allows residents to commute to Stockholm in about 30 minutes.

Because real estate prices in Uppsala have historically been lower compared to Stockholm, the capital of Sweden, the transportation infrastructure continues to make it attractive to live and do business in Uppsala today.

4.1.3. Sustainment: Acquisition, Relocation, or Early Initial Public Offering

Several interviewees emphasized that Uppsala's EE has gradually diversified and become less dependent on the life sciences industry, although the cluster is still critical to the region's success. The continued importance of life sciences in Uppsala is supported by the number of people employed in the cluster. In 2012, Uppsala had the highest density of employees in the life sciences industry in Sweden (see Figure 12). In 2016, life science cluster companies in Uppsala employed more than 5,300 people (full-time equivalents), making it the most labor-intensive sector in Uppsala (City of Uppsala, 2015). For a list of the ten life science companies with the most employees in Uppsala, see Table 3.

Despite the importance of the regional life sciences cluster, some of the entrepreneurial success stories mentioned above did not originate in the life sciences industry, but in unrelated fields such as finance (Klarna) and ICT (Skype, MySQL). The entrepreneurs behind such successful indigenous ventures not only inspire the local Uppsala population to also engage in entrepreneurial activities, but according to several interviewees, often engage in other activities that later contribute to strengthening the local EE. These activities include further entrepreneurial endeavors, i.e., becoming

Table 1: Data Structure of Uppsala's Ecosystem Birth.

Uppsala's Entrepreneurial Ecosystem			
Phase: Birth			
Domain	**Sub-Domain**	**Theme**	**Exemplary Quotes: IDs** (see appendix 8)
Culture	Societal Norms	Uppsala's reputation as an academic town	*UE2:1*
Human Capital	Educational Institutions	Research breakthrough in life sciences at Uppsala University	*US1:1*
		Relocation of Pharmacia motivated by university-industry R&D collaboration	*US3:1*
		Professor's privilege as hybrid entrepreneurship enablement	*UU1:1; US3:2*
		University-industry R&D collaboration attracts MNEs	*US3:3; UG1:1; UU1:2*
	Labour	Spin-off activities by former MNE employees	*UI1:1*
Markets	Networks	Dissolvement of Pharmacia attracting MNEs	*UE2:2*

Table 2: Data Structure of Uppsala's Ecosystem Growth.

Uppsala's Entrepreneurial Ecosystem			
Phase: Growth			
Domain	**Sub-Domain**	**Theme**	**Exemplary Quotes: IDs** (see appendix 8)
Markets	Networks	Uppsala's life sciences cluster continues to grow	*UE2:3; US2:1*
Policy	Leadership	Positive sentiment towards entrepreneurial activity among government officials	*UG1:2*
	Government	Significant public investment in entrepreneurial support infrastructure	*UG1:3*
Culture	Societal Norms	Close collaboration between relevant actors within Uppsala's EE	*US2:2; US2:3*
	Success Stories	Successful start-ups from Uppsala inspire nascent entrepreneurs	*UG1:5; UE2:4*
Human Capital	Educational Institutions	University-industry R&D collaboration as an important contributor to Uppsala's attractiveness	*UI1:2; UG1:4*
	Labour	Availability of well-educated human capital	*US1:2; US2:4*
Supports	Infrastructure	Proximity to Stockholm and Arlanda Airport contributes to Uppsala's attractiveness	*US1:3; UE2:5; UG1:6*

Table 3: Employment by Top 10 Life Science Companies in Uppsala in 2016. Source: STUNS Life Science (2016).

Company	**Employees** (2016)	**Activities in Uppsala**
GE Healthcare Bioscience	1,177	R&D, Manufacturing, M&S
Fresenius Kabi	947	R&D, Manufacturing, M&S
Thermo Fisher Scientific	544	R&D, Manufacturing, M&S
Galderma	469	R&D, Manufacturing, M&S
Recipharm Uppsala	203	R&D, Manufacturing
Quintiles	174	Service
J&J (AMO)	121	Manufacturing
GEMS PET Systems	105	R&D, Manufacturing, M&S
Orexo	102	R&D
Oasmia	75	R&D, Manufacturing, M&S
Total	**3,917**	

serial entrepreneurs, taking advisory positions in local start-ups, or becoming investors in regional businesses. In the interviews, these processes, combined with the collaborative mentality prevalent in the region, were described as a kind of pay-it-forward culture that has gradually developed throughout Uppsala's EE.

However, there are certain trends that pose a challenge to maintaining the local EE in Uppsala. It was highlighted

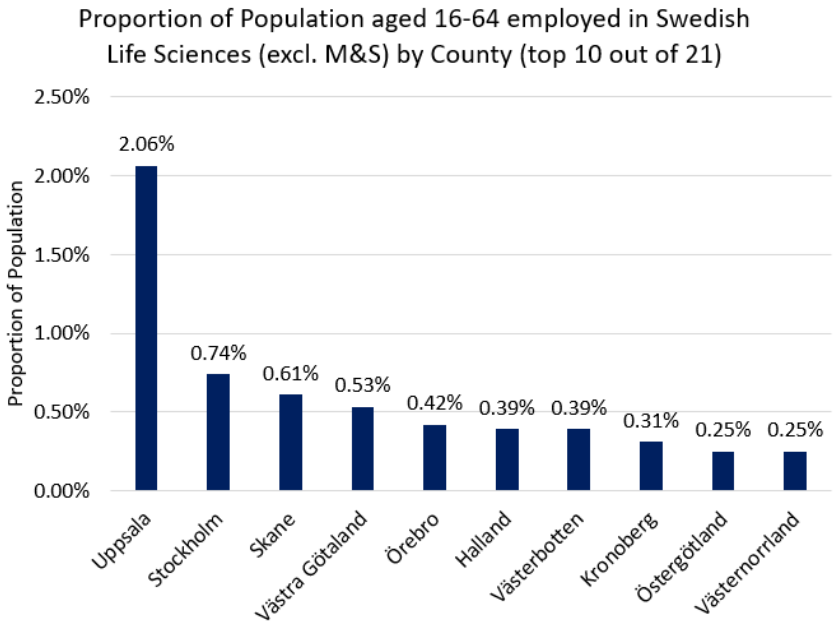


Figure 12: Proportion of Population Employed in Swedish Life Sciences in 2012. Source: adapted from Sandström (2014, p. 117).

in the interviews that despite some successful startups that have emerged from the EE in Uppsala, these companies tend to move either to Stockholm or to other international metropolises during their scale-up phase. This is at least partly due to the lack of availability of financial capital for regional scale-up companies, which limits their growth potential within Uppsala's EE. This was also noted by Eliasson and Eliasson (2006) when they analyzed the aftermath of Pharmacia's dissolution. Therefore, Uppsala may initially benefit from promising startups, but loses them at some point in their growth phase. Moreover, the lack of growth capital also leads to an increase in acquisition activity by local multinational subsidiaries (Eliasson & Eliasson, 2006). However, such acquisitions prevent Uppsala's EE from becoming less dependent on foreign multinationals and strengthen the domestic part of the local EE, as was also pointed out in the interviews. On the contrary, as Eliasson and Eliasson (2006) point out, a lack of qualified venture capitalists with sufficient industry knowledge increases the risk of regional activities retreating to other locations within or outside the country of origin, as well as missing out on technologies with significant economic potential. To date, Uppsala has been successful in retaining local MNE subsidiaries primarily due to university-industry collaboration, availability of skilled human capital, and relatively low wages (Eliasson & Eliasson, 2006). However, the continued dependence on foreign MNE subsidiaries, although less pronounced than at the beginning of the EE, poses a risk to Uppsala's economy, as one interviewee pointed out:

> "*Our main issue is that they [foreign-owned MNE subsidiaries] are all owned by American risk capital firms, so we are just a little dot. If they say, we do not want this little dot in this little town anymore because it is far away up north, then our whole ecosystem probably falls.*"

However, there is a notable alternative to VC investing in Sweden in the form of a small-cap venture exchange called First North (see Carpentier et al., 2010). This specialized exchange allows young companies to go public at an early stage and raise money through an initial public offering (IPO). While some regional companies make use of this alternative, according to several interviewees, Uppsala-based startups in their growth phase still largely aim for a takeover by a multinational company or relocate to a metropolis.

4.1.4. Summary

In summary, the analysis of Uppsala's EE shows that an EE depends on different actors and can emerge from processes that are not directly aimed at promoting local entrepreneurship. The emergence of Uppsala's EE can be attributed to Uppsala University's strength in life science research. This strength has led large life science companies, particularly Pharmacia, to locate in Uppsala and collaborate with the university on research and development. Through the creation of spin-off companies, these processes have led to the emergence of a strong life science cluster that continues to play an important role today, distinguishing the ecosystem from that of nearby Stockholm.

Through mergers and acquisitions, the subsequent breakup of Pharmacia has attracted other multinational companies operating in Uppsala, predominantly R&D. In addition, former employees, usually highly qualified and experienced, have contributed to Uppsala's EE growth by establishing spin-off companies. Government initiatives to support startups in Uppsala, both financial and non-financial, have enabled the emergence of such companies.

Table 4: Data Structure of Uppsala's Ecosystem Sustainment.

Uppsala's Entrepreneurial Ecosystem			
Phase: Sustainment			
Domain	**Sub-Domain**	**Theme**	**Exemplary Quotes: IDs** (see appendix 8)
Markets	Networks	Gradual diversification of Uppsala's EE	*US3:4; US1:4*
Culture	Societal Norms	Pay-It-Forward type culture fosters indigenous start-ups	*US1:5; UE1:1; UI1:3*
Finance	Financial Capital	Lack of growth capital in Uppsala	*UE2:6; UI1:4*
		Possibility of an early IPO at a small-cap venture exchange called First North	*UE2:7*
		Lack of growth capital leads to relocation of indigenous start-ups in the scale-up phase	*UG1:7; UI1:5; UI1:6*
		Lack of growth capital leads to increased acquisition activities by local MNE subsidiaries	*US1:6; US3:5*

Uppsala has also benefited from its size and spatial context, which has resulted in a high density of highly skilled human capital, geographic proximity of relevant actors within the EE, and proximity to relevant infrastructure. This has facilitated collaboration, for which Uppsala's inherent pay-it-forward culture has played an important role by minimizing competition within the system.

The case of Uppsala furthermore sheds light on problems that other second-tier regions may also face. Because Uppsala has a strong academic past, the city's reputation depends largely on its two local universities. As a result, Uppsala is perceived primarily as a university town, which to some extent discourages people from considering the region for their careers. In addition, the size of the region, combined with its proximity to Stockholm, makes it easier for both companies and individuals to relocate, making it difficult to retain companies and talent attracted by the allure of such a global metropolis. Another contributing factor is that Uppsala does not have access to growth capital, which would be important for startups looking to scale up. As a result, it is difficult for Uppsala to both convince thriving companies that have originated in Uppsala to stay in the region rather than move to a larger city, and to convince non-local startups and aspiring entrepreneurs to choose Uppsala as a location for their entrepreneurial endeavors. Figure 13 provides an overview of the development of Uppsala's EE.

4.2. Analysis of Galway's Entrepreneurial Ecosystem

Various aspects of Galway's medical technology cluster have been analyzed in the academic literature (Evers & Giblin, 2017; Giblin & Ryan, 2012; P. Ryan & Giblin, 2012; P. Ryan et al., 2020, 2021). The analysis of Galway's EE therefore combines secondary data with information obtained through qualitative interviews with actors within the Galway EE. The context of the Galway case is summarized in Figure 14.

4.2.1. Birth: Monetary Incentives Attract MNEs to Galway

As highlighted both in the literature and by interviewees, Galway's EE has grown rather organically (see, e.g., Evers and Giblin, 2017, p. 112). While the first foreign-owned medical technology company in Galway was established in Galway in 1973 (P. Ryan et al., 2021), the first pivotal event occurred in 1982 when CR Bard (acquired by Medtronic in 1999, hereafter referred to as Medtronic for simplicity) established a manufacturing facility in Galway (P. Ryan et al., 2021; P. Ryan et al., 2020; Evers and Giblin, 2017, p. 112; P. Ryan and Giblin, 2012). After CR Bard filed the first medical technology patent in the region in 1991 (P. Ryan et al., 2020), Boston Scientific, the second major MNE in the region, established a manufacturing site in 1994 (P. Ryan et al., 2021; Evers and Giblin, 2017, p. 112; P. Ryan and Giblin, 2012; Giblin and Ryan, 2012). Both multinationals were attracted to the region by IDA Ireland initiatives such as a low corporate tax rate, grants, and research incentives (P. Ryan et al., 2021; Evers and Giblin, 2017, p. 110; P. Ryan and Giblin, 2012; Giblin and Ryan, 2012), access to European markets, English-speaking labor, and a relatively low cost base (P. Ryan & Giblin, 2012). With the aforementioned incentives, IDA Ireland aimed to attract foreign direct investment (FDI) to benefit from the knowledge spillover from abroad (Evers & Giblin, 2017, p. 111; see also Figure 15).

Some interviewees have also linked IDA Ireland's efforts to attract foreign multinationals to a series of events related to Digital Equipment Corporation, a major player in the minicomputer industry. In 1971, the company opened an assembly plant in Galway (Van Egeraat & Jacobson, 2004). The assembly plant closed in 1993, and although Digital Equipment Corporation maintained a software development function in Galway (Van Egeraat & Jacobson, 2004), the plant's closure resulted in the loss of over 700 jobs in the region (Coughlan, 2017). As some interviewees pointed out, and although some of those laid off later became entrepreneurs in the region out of necessity, a local task force in collaboration with IDA Ireland worked to attract large new employers to the region after the closure (Coughlan, 2017), eventually succeeding with the arrival of Boston Scientific. Boston Scientific, which was also attracted by Medtronic's successful operations in Galway, then took over the empty premises previously occupied by Digital Equipment Corporation.

	Birth	Growth	Sustainment
Pull Factors	• Breakthrough research in biotechnology at Uppsala University • Relocation of Pharmacia to Uppsala	• Establishment of MNE subsidiaries within life sciences • Emergence of government-led EE support system	• Expansion of support system • Quality of life combined with proximity to Stockholm and Arlanda Airport
		Success stories and entrepreneurial recycling	
Policy	• University-industry R&D collaboration • Professor's privilege facilitates research commercialisation	• Publicly-financed innovation support organisations • Improvement of EE support processes (e.g. issuing of permits and licenses)	• Infrastructure development projects • Location marketing to initiate change of Uppsala's perception as mere university town
Finance	Government-led via university funding and private sources primarily coming from Pharmacia	Government-led, emergence of private investment	• Government-led with continued private investment • Lack of growth capital
		• Possibility of early IPO via small-cap venture exchange • Acquisitions by MNEs	
Culture	Strong academic reputation, entrepreneurial spirit rather limited	Entrepreneurship starting to gain traction	Entrepreneurship as a valid career path
		• Success stories contribute to continuous growth of interest in entrepreneurship, especially among local university students • Pay-It-Forward type culture emerges early on within the EE • Collaboration rather than competition between EE actors	
Support	University-based research commercialisation support system	Emergence of public and private support organisations	Extensive collaboration between different support organisations within the EE
		Increase of university-led research commercialisation efforts	
Human Capital	R&D crucial to emergence of entrepreneurial endeavours	Some competition for talent with Stockholm	
	• Influx of students and researchers • University researchers may engage in hybrid entrepreneurship		
	Pharmacia's gradual dissolvement leaves highly skilled human capital with industry experience		Economic potential in Uppsala starts to get noticed
Markets	• Clear focus on life sciences industry driven by academic research and presence of Pharmacia • Life sciences as a globalised industry	• Life sciences cluster continues to grow • MNEs open subsidiaries • Emergence of indigenous firms with a rather regional focus in the beginning	• Life sciences cluster still crucial, but start of EE diversification in industries that also benefit from university-industry R&D collaboration • Emergence of indigenous firms with a born global internationalisation approach

Figure 13: Development of Uppsala's Entrepreneurial Ecosystem. Source: own illustration, format adapted from Mack and Mayer (2016).

While the two multinationals initially established only manufacturing facilities in Galway, Medtronic established an R&D center in 1996, followed by Boston Scientific's product development center in 1997 (P. Ryan et al., 2021). The establishment of innovation-oriented facilities was critical both to the survival of the multinationals' subsidiaries and to the development of Galway's EE as a whole, as one researcher interviewed pointed out. While early local start-up activity in the 1980s and 1990s focused predominantly on supplying the two multinationals (Evers and Giblin, 2017, p. 114; P. Ryan and Giblin, 2012), it was indigenous firms founded by former employees of the multinationals that designed and developed their own ideas in areas unrelated to the industry (P. Ryan & Giblin, 2012; P. Ryan et al., 2021). The former Medtronic and Boston Scientific employees benefited from the connections they made and the management skills they developed while working for the MNEs (P. Ryan et al., 2021; Evers and Giblin, 2017, p. 118; P. Ryan and Giblin, 2012). These connections and skills, along with the international exposure, quality compliance experience, and credibility they built as business professionals, enabled them to identify business opportunities leading to spin-off activities in Galway (P. Ryan et al., 2021; Evers and Giblin, 2017, p. 118; P. Ryan and Giblin, 2012). Because the medical technology market is global, international connections were critical to commercial success (Evers and Giblin, 2017, p. 118; Giblin and Ryan, 2012). However, spin-off activities were not limited to former MNE employees. As mentioned by interviewees, researchers from

Context of Galway's Entrepreneurial Ecosystem	
Location	West of Ireland (~210 km west of Dublin, ~215 km west of Dublin Airport)
Population	City of Galway: ~80,000 | County Galway: ~258,000
EE Size	198 start-ups in County Galway listed on Crunchbase[1]
Funding	USD 448.2m (122 funding rounds) since 2000 listed on Crunchbase[1]
Industry Cluster	Medical Technology
Success Stories[2]	Altocloud (ICT) | Creganna (MedTech) | Aerogen (MedTech) | Planet (FinTech)
Universities	National University of Ireland, Galway (NUIG) | Galway-Mayo Institute of Technology (GMIT)
Support Organisations[2]	GMIT Innovation Hub | NUIG Innovation Office | Galway City Innovation District | SCCUL Enterprises | WestBIC
Incubators / Accelerators[2]	EI New Frontiers at GMIT | StartLab Galway | BioInnovate at NUIG
Coworking Spaces[2]	PorterShed | Galway Technology Centre | SuperPixel Labs
Financial Capital[2]	Enterprise Ireland | Western Development Commission | NDRC | EASME
Notable MNEs[2]	Medtronic | Boston Scientific | Electronic Arts

[1] Crunchbase is increasingly being used as a database for scientific research in economics and management (Dalle et al., 2017), which is why its data is used to provide an overview of Galway's EE. Access date: 20 July 2020.
[2] Main actors according to interviewees and secondary data; therefore the list is non-exhaustive.

Figure 14: Case Context of Galway. Source: own illustration.

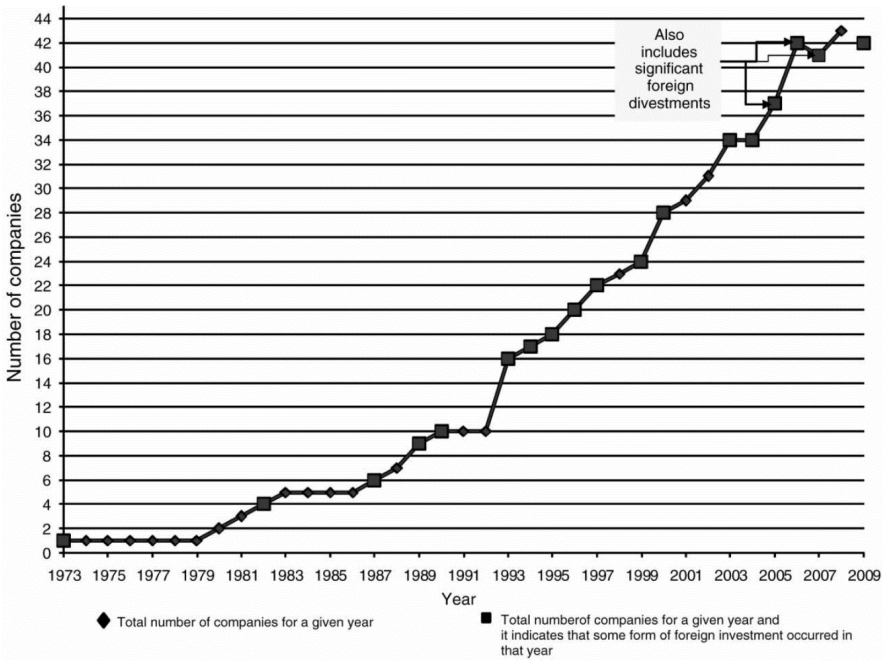


Figure 15: Medical Technology-Related Companies in Galway from 1973-2009. Source: Giblin and Ryan (2012, p. 251); the typing error in the figure's legend ("total number of") has been noted.

local universities also began to participate in the start-up processes. This then accelerated during the growth phase of the EE when local universities began to establish specialized research centers that began to attract highly qualified students and researchers to the Galway region.

Table 5: Data Structure of Galway's Ecosystem Birth.

Galway's Entrepreneurial Ecosystem			
Phase: Birth			
Domain	**Sub-Domain**	**Theme**	**Exemplary Quotes: IDs** (see appendix 8)
Policy	Government	Monetary incentives to attract MNEs	*GU1:1; GU1:2*
		Business-friendly regulatory environment	*GE1:1*
Human Capital	Educational Institutions	Local universities provide skilled human capital	*GS1:1*
		R&D activities at local universities lead to spin-off activities by students and researchers	*GI1:1; GS1:2*
	Labour	R&D activities within local MNE subsidiaries lead to spin-off activities in related industries	*GI1:2; GU1:3; GS1:3*
Markets	Networks	Influx of MNEs from the MedTech field and their shift from production to R&D enable the development of an industry cluster in Galway	*GG1:1; GG1:2*

4.2.2. Growth: NUI Galway Adapts to Emerging MedTech Cluster

While the main local university, NUI Galway, did not initially function as an anchor within the EE, it has continually responded to economic developments in Galway. Since the establishment of an engineering department in 1980 (P. Ryan et al., 2021), NUI Galway has become a major research partner and a provider of specialized human capital. In 1998, the university introduced a biomedical engineering degree program (Giblin & Ryan, 2012; P. Ryan et al., 2021), and in 1999, the National Centre for Biomedical Engineering Science (NCBES) was established to foster university-industry collaboration (Giblin & Ryan, 2012; P. Ryan et al., 2021). In 2003, NUI Galway established the Regenerative Medicine Institute (REMEDI) to further enhance public-private collaboration, and in 2009, a dedicated postgraduate diploma in medical device science was introduced (Giblin & Ryan, 2012; P. Ryan et al., 2021). Shortly thereafter, in 2010, the university initiated the BioInnovate Ireland training program aimed at producing indigenous medical device startups (P. Ryan et al., 2021). Since then, NUI Galway has continued to increase its importance to the local EE, for example, by introducing a master's degree program in biomedical engineering in 2013 and establishing the Irish Centre for Cell Manufacturing (ICCM) and the Centre for Research in Medical Devices (CURAM) in 2014 (P. Ryan et al., 2020, 2021). Overall, the two local universities, NUI Galway, and the Galway-Mayo Institute of Technology (GMIT), were important sources of skilled human capital and frequently collaborated with both foreign and local companies, for example, through research, student placements, or the provision of office space (Giblin & Ryan, 2012; P. Ryan & Giblin, 2012). These activities were not limited to the medical technology industry. As noted by several interviewees, both NUIG and GMIT contributed to the development and subsequent diversification of the EE by attracting financial resources, establishing research centers, attracting graduate students and high-level researchers, and providing startups with the infrastructure they needed to develop.

Following the establishment of subsidiaries by CR Bard / Medtronic and Boston Scientific, Galway's EE has developed rather organically and naturally. While the two multinationals were initially attracted by incentives from IDA Ireland aimed at attracting and retaining FDI, the availability of skilled human capital, research collaboration with local universities, and a growing medical technology cluster have contributed to the growth of the EE. Former employees of multinationals have started their own businesses, first in related and later in unrelated industries, which has enabled the gradual diversification of the region and the growing independence from multinationals. This has also been enabled by several local support organizations that have formed, as well as other relevant actors, such as investors, who began to get involved in the growth phase of Galway's EE. However, as several interviewees emphasized, the growth of the local support system for startups and future entrepreneurs has been driven more by initiatives of key private individuals than by larger local government initiatives. Nevertheless, the importance of public funding that contributed to the growth of EE should not be underestimated, particularly the funding provided by actors such as Enterprise Ireland, IDA Ireland, and the Local Development Office, especially in the early stages of venture development.

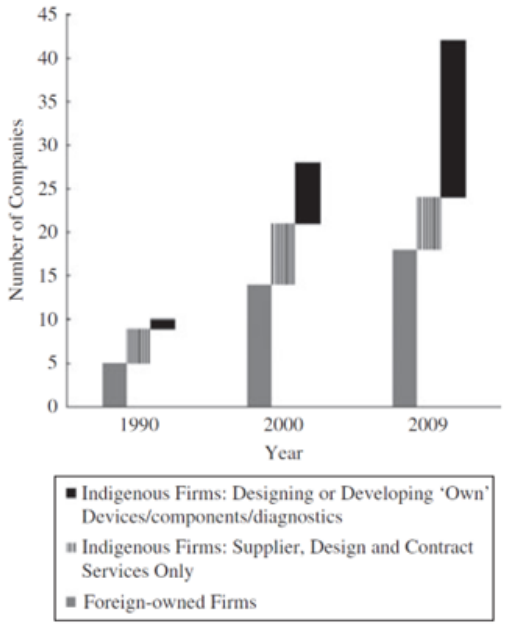
4.2.3. Sustainment: Acquisition as Favored Exit Strategy

Because an EE is made up of diverse actors engaged in innovation, knowledge transfer, research, and other entrepreneurial activities, multinational companies alone, which in the case of Galway initially served as anchor organizations and surrogate universities, were not sufficient to build an active ecosystem (Evers & Giblin, 2017, p. 120). The EE in Galway, while initially focused on a subset of the medical technology industry, has diversified over time, preventing phenomena such as technological lock-in, cluster hollowing-out, and overspecialization (Evers and Giblin, 2017, pp. 117-120; Clancy et al., 2013). Diversification and capability development within Galway's EE have also contributed to growing independence from local MNE subsidiaries (P. Ryan and Giblin, 2012; P. Ryan et al., 2020; see also Figure 16) .

In particular, domestic global companies have developed

Table 6: Data Structure of Galway's Ecosystem Growth.

Galway's Entrepreneurial Ecosystem			
Phase: Growth			
Domain	**Sub-Domain**	**Theme**	**Exemplary Quotes: IDs** (see appendix 8)
Human Capital	Educational Institutions	University-industry R&D collaboration as an important contributor to Galway's attractiveness	*GG1:3; GU1:4*
	Labour	Preparation of local human capital for the local industry by universities	*GS1:6*
Supports	Infrastructure & NGOs	Continuous development of support infrastructure by a group of volunteers and their non-governmental institutions	*GI1:3; GS1:4*
Policy	Government	Higher education funding as crucial investment in Galway's EE	*GE1:2; GS1:5*



Note:
(i) "For the purposes of this graph, the design or development of 'own' devices, components or diagnostics by indigenous firms refers to those firms that are developing their own ideas as opposed to being solely a supplier or contractor. The output may not necessarily be sold or aimed to be sold in the marketplace under the company's own name.

Figure 16: Activity of Indigenous Ventures in Galway's Medical Cluster for Given Years. Source: P. Ryan and Giblin (2012, p. 1334).

products and services that target market niches and use novel technologies (Evers & Giblin, 2017, p. 119) to avoid competition with local MNE subsidiaries. The creation of such companies and the frequent mergers, acquisitions, and management buyouts underscore the dynamic nature of the ecosystem (P. Ryan & Giblin, 2012). Evers and Giblin (2017, pp. 122-123) also find that the founders of born global firms in Galway are more likely to seek the acquisition of developed and commercialized technology by an MNE than to grow independently. According to several interviewees, this is at least partly due to the difficulty of raising capital in Galway, given the lack of local growth capital providers. In cases where ventures do want to grow independently, there is a tendency for these companies to relocate to metropolitan areas during their scale-up phase. The founders of local startups that nevertheless decide to stay in the region tend to later either become serial entrepreneurs, participate in mentoring programs, or engage in investment activities (Evers and Giblin, 2017, pp. 115-116; P. Ryan and Giblin, 2012). In the literature, such processes are referred to as entrepreneurial recycling (Mason & Harrison, 2006).

4.2.4. Summary

The Galway case shows that an EE can emerge organically and as a byproduct of more disjointed structured policy approaches (Evers & Giblin, 2017). The evolution of Galway's EE demonstrates that new businesses in the start-up and early growth stages require structurally embedded support and leadership programs to ensure their long-term survival, which in turn directly impacts local employment security (Evers & Giblin, 2017). However, in addition to public and

Table 7: Data Structure of Galway's Ecosystem Sustainment.

Galway's Entrepreneurial Ecosystem			
Phase: Sustainment			
Domain	**Sub-Domain**	**Theme**	**Exemplary Quotes: IDs** (see appendix 8)
Markets	Networks	Gradual diversification of Galway's EE	*GG1:4; GS1:7*
Culture	Success Stories	Local success stories inspire others to start entrepreneurial ventures	*GU1:5; GU1:6*
	Societal Norms	Entrepreneurial recycling strengthens the local EE	*GU1:7; GU1:8; GI1:4*
Finance	Financial Capital	Lack of growth capital in Galway	*GI1:5; GG1:5*
		Lack of growth capital leads to relocation of indigenous start-ups in the scale-up phase	*GE1:3*
		Lack of growth capital leads to increased acquisition activities by local MNE subsidiaries	*GG1:6; GI1:6; GU1:9; GG1:7*

private support from volunteers and nonprofits, and due to the importance of spinoff activities in Galway, corporate-level initiatives are also needed to foster innovation and entrepreneurship in the region, such as management and commercialization training (Evers & Giblin, 2017). Furthermore, when considering the role of NUI Galway and the Galway-Mayo Institute of Technology in Galway's EE, university-industry collaboration in research and development and the availability of specialized human capital are critical to entrepreneurial activities, which can be ensured through sufficient higher education training and research funding.

When a government invests in the local EE, it must also ensure that technology developed by indigenous startups ultimately leads to the creation of a new business in the region that creates local jobs and reduces dependence on multinational enterprises (Evers & Giblin, 2017, pp. 122-123). In Galway, researchers have observed that after developing a new technology, founders are more likely to seek an acquisition by an MNE than to develop their businesses independently (Evers & Giblin, 2017, p. 122). However, such acquisitions do not help an EE becoming less dependent on MNEs. Therefore, a government should provide incentives to local entrepreneurs to build independent indigenous ventures rather than selling technology early. Policymakers also need to take a holistic approach. Instead of focusing only on jobs created by existing companies at a given time, job creation by serial entrepreneurs must also be considered (Evers & Giblin, 2017, p. 122). Therefore, policies should take a multiyear perspective and focus on human capital development over a longer period.

Interviewees also pointed out the weaknesses of Galway's EE. First, the geographic distance to a major international airport and the underdeveloped transit infrastructure are a major disadvantage for entrepreneurs focused on international markets, as the long travel times both domestically and internationally require large time and financial resources. Second, the lack of financial growth capital in the region leads to increased M&A activity by MNEs and relocation in the scale-up phase, which prevents the emergence of large domestic companies. Third, entrepreneurial support in Galway driven by individuals and organizations is constrained, at least in part, by other stakeholder commitments, as local EE support is dependent on volunteer labor. Fourth, while there are several public policies and strategies aimed at promoting economic development in the Irish regions, some interviewees feel that there is too much focus on Dublin, limiting public funding and thus the development potential of the Galway region.

The Galway case shows a development of technological and intangible capabilities, such as expertise in management and networking at the international level (P. Ryan & Giblin, 2012). Awareness of their own strengths, sufficient foresight, and adequate knowledge were critical for local decision makers, private volunteers, and nonprofit organizations to enable the emergence and growth of Galway's EE (P. Ryan & Giblin, 2012), illustrating the complexity, long-term nature, holistic nature, and spatial contextuality of such processes. Figure 17 provides an overview of the evolution of Galway's EE.

4.3. Comparative Analysis of Uppsala and Galway

While there are some similarities in the analyzes of the EEs in Uppsala and Galway, there are also relevant differences that have policy implications for other second-tier European regions seeking to establish and grow an EE.

In terms of commonalities, both EEs initially revolved around a specific industry cluster and later diversified into related and unrelated industries, mitigating cluster risks. The two clusters, medical technology, and life sciences / biotechnology, both primarily serve international markets, so export-enhancing spatial characteristics have been important in both regions, and emerging new firms in these industries often identify as "born globals." In addition, both regions demonstrated the importance of anchor organizations in general, even if the nature of these organizations differed. In both regions, however, multinationals and large corporations played an important role in the development of the local EE, attracted skilled workers, and facilitated the creation of spin-off companies by former employees. Both regions also benefited from a quality of life perceived to be better compared to their respective capitals, which provided an incentive for skilled workers to stay and build a livelihood, e.g., by starting

	Birth	Growth	Sustainment
Pull Factors	• IDA Ireland incentives for MNE subsidiaries • Presence of CR Bard and Boston Scientific	• Relocation of MNEs R&D activities to Galway • Emerging MedTech cluster	• Expansion of support system • University-industry collaboration
Policy	Tax incentives and research grants to attract MNEs	• Continued investment in higher education and research • Low corporate tax rate and general facilitation of doing business in Ireland	
		Local government rather passively involved in EE	EE starts to get more attention by local government
Finance	Mainly driven from the private side due to attraction of FDI by providing monetary incentives to MNEs	Driven by FDI	Lack of growth capital
		• Early-stage capital provided to new ventures from both public and private funds • Acquisitions by MNEs as entrepreneurs' primary exit strategy	
Culture	Large MNEs establishing subsidiaries in Galway provide job opportunities for local workforce and are seen as key to economic welfare in the region	First successful indigenous ventures motivate other skilled people to start own companies	Entrepreneurship is seen as crucial aspect of regional economic development
Support	Only a few entrepreneurial support systems in place which are mainly driven by private individuals and their organisations	Support system starts growing, mainly driven by private individuals and their organisations with moral support of local government	Local government starts focusing more on entrepreneurial activity, main drivers of support network remain private individuals and their organisations
Human Capital	Relocation of MNEs production facilities indicate availability of "productive" workforce	Specialisation of NUI Galway to provide local economy with highly skilled human capital in MedTech areas	Industry-aligned specialised research and teaching at local higher education institutions facilitate recruitment of skilled human capital
Markets	• Production focus within MedTech industry • Little R&D activity by local MNE subsidiaries • MedTech as a globalised industry	• Growth of local MedTech cluster • R&D activities gradually expand within MNEs • Emergence of indigenous firms focusing on supplying local MNE subsidiaries	• Continued diversification of EE and decreasing importance of foreign-owned MNE subsidiaries • Emerging indigenous firms with a born global internationalisation approach

Figure 17: Development of Galway's Entrepreneurial Ecosystem. Source: own illustration, format adapted from Mack and Mayer (2016).

their own businesses, despite job losses in the subsidiaries of multinational enterprises. Moreover, the presence of large companies accelerated the development of the respective industrial cluster and contributed to employment prospects in the regions. In addition, university-industry partnerships, especially in R&D activities, proved to be crucial for the ecosystems. Regarding the sustainment of the EEs, both regions have struggled to attract growth capital, resulting in indigenous companies either being acquired early by larger companies or relocated after reaching the scale-up stage. In Uppsala, however, small companies can go public early through a dedicated stock exchange, which provides an alternative to raising venture capital.

A notable difference in the formation phase of the ecosystems is that companies were attracted to Uppsala by significant scientific findings from Uppsala University, while Galway attracted companies by monetary incentives provided by a government agency, IDA Ireland. This is also the reason why Uppsala companies moved R&D activities to the region from the beginning, while Galway companies first moved mainly production and only later R&D activities. There are also differences in the companies that were relevant to the emergence of each EE. Pharmacia was a domestic company, while both CR Bard and Boston Scientific were large U.S. corporations and thus foreign companies.

The later development of the two EEs revealed further differences. The EE in Galway grew more organically, with multinationals acting as anchors and incubators for new businesses. In Uppsala, growth was due to a combination of government initiatives and processes within large companies that also led to the creation of spin-off companies. The importance of local universities in the different stages of development also varied across the regions. In Uppsala, Uppsala University was a key player in the early stages of the ecosystem (action), while NUI Galway gradually gained importance by specializing in areas relevant to the local industry cluster

(reaction).

Even if both regions are considered second tier, there are relevant differences in spatial characteristics that influence both EEs. Uppsala's proximity to Stockholm and Arlanda Airport has a predominantly positive impact on the ecosystem by facilitating internationalization and increasing the available labor pool, although there is some competition for skilled labor between Uppsala and Stockholm. Galway's geographic location, on the other hand, limits competition for human capital with Dublin, the Irish capital. However, due to its distance from a major airport and the lack of a high-speed train connection, both domestic and international travel is longer compared to Uppsala, which negatively impacts Galway's attractiveness to startups whose employees need to travel frequently. An overview of the similarities and differences in the development processes that characterize the EEs in Uppsala and Galway can be found in Figure 18.

After analyzing and comparing the EEs in Uppsala and Galway, and by combining the findings with the literature review in section 2, the section 5 addresses the resulting policy implications for second-tier European regions seeking to establish and grow a local EE.

5. Policy Implications

In an empirical analysis of the quality of EEs in the U.S., Vedula and Kim (2019) showed that the quality of an EE affects the chances of firm survival. Nascent entrepreneurs benefit significantly from high-quality EEs, while serial entrepreneurs are less dependent on the quality of an EE. Vedula and Kim's analysis implies that nascent entrepreneurs in particular need to be strategic in their location decisions to maximize the chances of firm survival. Even for entrepreneurs already located in a lower-quality EE, moving to a higher-quality EE brings additional challenges, such as acquiring knowledge about the new market, building a good reputation, creating new networks, and mobilizing resources (Vedula & Kim, 2019). Therefore, location decisions in the early stages of venture development are particularly important for nascent entrepreneurs. This underscores the need to create and maintain a high-quality EE in regions that want to leverage entrepreneurship for regional economic growth.

As highlighted in earlier sections, policymakers must keep in mind that not all entrepreneurial activity contributes positively to a region's economic development (Acs et al., 2016; Baumol, 1990; Mason & Brown, 2014; Nightingale & Coad, 2014). While entrepreneurship in tradable sectors increases the likelihood of positive local spillover effects that benefit people not directly involved in entrepreneurial activities, entrepreneurship in non-tradable sectors shows a weaker impact on local economic growth and its effects, such as poverty reduction (N. Lee & Rodríguez-Pose, 2021). However, as Aparicio et al. (2020) have noted, innovation- and opportunity-driven entrepreneurship has the potential to contribute to social inclusion and inclusive growth, including in vulnerable communities.

Policies should be embedded in a larger framework aimed at encouraging local entrepreneurial activity. An EE, as explained in section 2.1, consists of several actors that influence its quality. Government actions therefore represent a part of the larger ecosystem, which underscores the need to synchronize government actions with incentives and support from other actors (Spigel, 2017). As Acs et al. (2016) point out, certain policies that can be highly effective in promoting local entrepreneurial activity may not be immediately recognizable as such, such as education policies or social security (Acs et al., 2016). Therefore, policymakers should take a holistic and systems-based approach to policy development and implementation. Policymakers should also ensure that policies are dynamic, as the ecosystem is constantly evolving and the sources of incentives for entrepreneurial activity change over time.

Combining the literature review on EEs in section 2 and the analysis of EEs in Uppsala and Galway in section 4 identified policy approaches that second-tier regions have used to establish and promote local EEs. Although spatial differences and a region's individual context must be considered, the following policy approaches can help second-tier regions develop a local policy framework and establish support systems that help create, grow, and sustain a thriving EE. The implications are divided according to the stages of development of an EE, which were presented in section 2.1 (Mack & Mayer, 2016). The decline phase was excluded because the policy implications for a declining EE are similar to those of the previous phases and are aimed at either rebirth or renewed growth. Figure 19 provides a brief overview of the key findings relevant for regional decision makers, while the following subsections provide a more detailed elaboration of the policy approaches and their respective implications.

5.1. Birth Phase

As the analysis of Uppsala and Galway, as well as the regions described in Appendix 4, has shown, second-tier regions tend to focus on a particular subset of an industry during the birth phase to build a solid knowledge base and specialize their activities. Given limited resources such as financial and human capital, it is likely necessary to build an EE around an existing industry cluster to benefit from comparative advantages (Giblin & Ryan, 2012). In addition, policymakers need to understand local strengths and weaknesses to recognize the region's capabilities and identify areas for smart specialization (P. Ryan & Giblin, 2012; Szerb et al., 2020). Furthermore, as Szerb et al. (2020) mention, policymakers should avoid focusing only on a region's strengths. Instead, mitigating weaknesses may be a necessary condition for realizing the full potential of regional strengths. The REDI presented in sections 2.4.3 and 2.5 can help regions in the EU identify such constraining pillars and design policies that alleviate existing bottlenecks (Szerb et al., 2020).

Next, policymakers should take a long-term perspective when attempting to establish an EE (P. Ryan & Giblin, 2012). Such ecosystems, as this article has repeatedly emphasized, go through various stages of development, and it likely takes

Theme	Similarities	
Industry Cluster	Both EEs revolve around a specific industry (Life Sciences/Biotech in Uppsala, MedTech in Galway)	
R&D Activities	R&D activity, within academia and/or within MNEs, was a prerequisite for the emergence of both EEs, also highlighting the importance of university-industry R&D collaboration	
Gradual Diversification	A gradual diversification is observable in both EEs	
Importance of MNEs	MNEs and large corporations, especially their R&D activities, played a major role in the EEs' development	
Regional Size	Both regions benefit from their limited size in terms of community building facilitation, a higher perceived quality of life by key actors, and cost benefits compared to the respective capital city	
Entrepreneurial Recycling	Entrepreneurial recycling and a Pay-It-Forward type culture are key phenomena in both EEs	
Success Stories	Regional success stories positively impact the emergence of nascent entrepreneurship in both EEs	
Growth Capital	There is a lack of growth capital in both EEs, leading to increased acquisition activities by MNEs and/or the relocation of scale-ups	
Theme	**Uppsala**	**Galway**
MNE Attraction	MNEs attracted to the region by research strength of local university	MNEs attracted to the region by monetary incentives provided by a government agency
Ownership of First Large Organisations	The first large organisation with a major impact on the emergence of the EE was native to Sweden	The first large organisations with major impacts on the emergence of the EE were foreign-owned
EE Development	EE development was spurred by both private and public initiatives	EE development was rather organic and driven by private initiatives
Anchor Organisations	Local university and Pharmacia acted as anchor organisations	MNEs acted as anchor organisations
Timing of MNEs R&D Activities	Large organisations and MNEs relocated R&D activities early on, production activities have not played a major role	Large organisations and MNEs relocated production activities early on and only later expanded into R&D activities
Spatial Context / Infrastructure	Proximity to Stockholm and Arlanda perceived as a positive factor for international travel and human capital availability, but some competition for talent with Stockholm	Lack of proximity to Dublin and international airport negative factor for international travel, skilled human capital is provided by local universities, and little competition for talent with Dublin

Figure 18: Comparison of Entrepreneurial Ecosystems in Uppsala and Galway. Source: own illustration.

several years for the effects of a local EE to become apparent when analyzing regional economic and financial indicators. Policy approaches should therefore be based on a long-term commitment to ensure the viability and resilience of the local EE, while also emphasizing the need for positive attitudes toward entrepreneurship among government officials and the public.

In the case of Galway, a low corporate tax rate, grant incentives, and research incentives motivated multinationals to establish subsidiaries in the region. The attraction of anchor organizations as a policy approach also worked in the cases of Portland, Boise, and Kansas City, although these case studies are not subject to the European context (see Appendix 4), again underscoring the need for contextual sensitivity. In any case, R&D activities in local affiliates of MNEs are particularly important to benefit from value-added spin-off activities. Moreover, trade policy enabled multinationals as well as emerging born global startups in Galway to do business outside of Ireland and provided international networking that is particularly important for small economies (P. Ryan & Giblin, 2012). Furthermore, the importance of networking in the birth phase, especially between existing and nascent entrepreneurs (Mack & Mayer, 2016), should not be underestimated.

The attraction of a large organization has also spurred entrepreneurial activity in Uppsala. The anchor in Uppsala's case, however, was its world-renowned university, which engaged in university-industry collaboration in research and development. Following the dissolution of Pharmacia, a number of highly qualified former employees who wished to remain in the region established their own entrepreneurial ventures. These processes were supported by a solid entrepreneurial support network, driven by both private and public actors. In addition, Uppsala encouraged spin-off activities and the commercialization of research results through the establishment of technology transfer offices at local universities to assist researchers in commercializing research. Although most researchers remain employed part-time at the university, support for research commercialization has produced hybrid entrepreneurs who work simultaneously in

	Development Stage		
Domain	Birth	Growth	Sustainment
Policy	Build on local strengths, mitigate weaknesses, and foster positive sentiment towards entrepreneurship among government officials	Facilitate new venture creation and research commercialisation via legal framework, promote the growing EE and local capabilities	Adapt the local policy framework and support mechanisms to cater for the needs of an increasingly diversified EE and local industry
Finance	Ensure early-stage capital availability for entrepreneurs from public and private funds	Incentivise private investment in early-stage and growth-stage ventures, start attracting skilled VCs and other growth capital providers to the region	Attract skilled growth capital providers to the region, ensure the accessibility of growth capital in the region to prevent high-growth start-ups' relocation or acquisition
Culture	Encourage networking between existing and nascent entrepreneurs	Establish collaboration-promoting physical spaces for relevant actors within the EE, look after serial entrepreneurs to foster entrepreneurial recycling	Use local success stories for promotional activities to encourage further regional entrepreneurial activity
Human Capital	Provide sufficient third-level funding to ensure the availability of skilled human capital and enable the commercialisation of research findings	Ensure alignment between higher education institutions' activities and the needs of the local industry cluster	Ensure continued third-level funding and universities' development to utilise strengths, mitigate weaknesses, and cater for the needs of the local industry
Supports	Establish support programmes helping nascent entrepreneurs develop necessary entrepreneurial skills	Develop physical infrastructure to ensure the attractiveness of the region to MNEs and born global start-ups, incl. regional, national and international transportation	Engage in conversations with entrepreneurs to determine and close gaps in the support system, ensure high-quality infrastructure
Markets	Focus on a specific industry subset to build a cluster, build an attractive business environment, attract MNEs, foster university-industry R&D collaboration	Enable the gradual diversification of the local industry via industrial, regional, science, and technology policy mechanisms	Incentivise the independent commercialisation and growth of regionally developed new technologies to avoid MNE overdependence

Figure 19: Overview of Policy Implications by Ecosystem Development Stage. Source: own illustration.

academia and business. The Swedish concept of professor's privilege has enabled these processes by preventing IP protection issues. A list of policy approaches to EE creation in second-tier European regions is shown in Figure 20.

While the policy approaches listed in Figure 20 contribute to the birth of an EE, the continuation of most of these policies is necessary during subsequent phases of development. In particular, sufficient third-level funding, university-industry collaboration in R&D, and commercialization of research results were also important for the growth and sustainment of the EEs in Uppsala and Galway.

5.2. Growth Phase

By investing in higher education, policymakers in Galway ensured that the medical technology cluster had access to skilled human capital. Similar processes around local universities, as discussed in Appendix 4, were also observed in Calgary (Spigel, 2017) and Chattanooga (Motoyama et al., 2016). In addition, in both Galway and Chattanooga (Motoyama et al., 2016), the gradual development of the local university to serve and collaborate with the emerging cluster of foreign and local companies was observed. Thus, not only companies but also universities specialized in activities that were aligned with the local cluster.

As the Galway and Uppsala cases have shown, policies that create incentives for current employees of local MNE subsidiaries, as well as for employees of local higher education institutions, to engage in spin-off activities can contribute to the growth of the EE (P. Ryan & Giblin, 2012). Such policies may relate, for example, to developing current production capacity, facilitating the creation of new ventures, training the necessary skills, regional, national, and international networking (P. Ryan & Giblin, 2012), and facilitating the commercialization of research results.

As Giblin and Ryan (2012) note, policy can also focus on attracting more FDI by promoting both the capacity and capabilities of emerging domestic firms, thereby creating incentives for foreign investors to further exploit these local capabilities. This also underscores the importance of geographically concentrated technological capabilities (P. Ryan & Giblin, 2012). Policies aimed at further developing these capabilities can therefore maximize the resulting external economies (P. Ryan & Giblin, 2012). Moreover, to ensure the resilience of the developing EE, policymakers should not focus exclusively on the emerging industry cluster, but should develop industrial, regional, science, and technology policy mechanisms that allow for technology and industry diversification (P. Ryan & Giblin, 2012). Gradual diversification of the EE mitigates cluster risks, so policy approaches aimed at diversification are also important in the sustainment phase.

Both Galway and Uppsala have benefited from their reputation in a particular industry. However, neither EE is particularly well known outside these industries. Therefore, using marketing techniques to increase awareness of the existing supportive ecosystem for entrepreneurs can help regions attract prospective and early-stage entrepreneurs, as well as other relevant stakeholders such as investors, to locate in their respective second-tier regions. In doing so, the advantages of the local EE should be highlighted. Second-tier re-

Birth	
Policy Approach	**Impact**
Focus on a specific subset of an industry	Build a solid knowledge base, specialise activities and create an impactful cluster
Comprehend local strengths and weaknesses	Identify the region's capabilities and identify areas of smart specialisation
Take a long-term approach to EE development and dynamically adjust policy measures	Prevent surrender if policy measures do not have an immediate (measurable) impact
Implement targeted (monetary) incentives and create an attractive environment to do business	Incentivise MNEs to establish local subsidiaries
Implement attractive trade policy and an infrastructure facilitating international trade	Incentivise MNEs to establish local subsidiaries and aid native born global firms' inception and growth
Encourage networking between existing and nascent entrepreneurs	Foster indigenous venture creation by local population
Establish and fund organisations dedicated to supporting existing and nascent entrepreneurs	
Encourage university-industry R&D collaboration via funding mechanisms	Attract MNEs and indigenous ventures to the region, encourage spin-off activities
Ensure sufficient third-level funding	Ensure the availability of highly skilled human capital and local R&D capabilities
Facilitate research commercialisation via technology transfer offices at universities and IP regulation	Encourage spin-off venture creation by researchers

Figure 20: Policy Approaches for Second-Tier Entrepreneurial Ecosystem Birth. Source: own illustration.

gions, such as those found in Uppsala and Galway, tend to benefit from dense networks, collaboration rather than competition, a sense of community, and a pay-it-forward culture. This also underscores the need for community development in second-tier EEs, as this can be a competitive advantage over EEs in superstar regions. A list of policy approaches for growing an EE in second-tier European regions is shown in Figure 21.

Overall, to grow a second-tier EE, having a support system in place that aids current and nascent entrepreneurs is necessary. Awareness thereof and collaboration within the ecosystem are also crucial to ensure the development of a sustainable competitive advantage. Related to this, and since a competitive advantage must be evaluated relative to competing EEs (Porter, 1985, pp. 1-30), competing regions should be analyzed thoroughly and continuously.

5.3. Sustainment Phase

As highlighted in Galway's EE analysis, sustaining an EE requires avoiding phenomena such as technological lock-in (see also, e.g., Maggioni, 2004). Thus, policies should focus on diversifying the ecosystem to mitigate cluster risks. Mack and Mayer (2016) suggest networking the local EE with other EEs at the national and international levels to avoid entering the decline phase. Seeking international networks and sharing ideas and experiences with other second-tier EEs was also highlighted by a government representative as a goal of Galway's economic policy.

In both Uppsala and Galway, a lack of local growth capital has been observed as a particular risk that can lead to stagnation and/or entry into the decline phase. This is also a challenge highlighted in a paper by Xu and Dobson (2019) on second-tier EEs. Due to their limited size, second-tier regions are not necessarily on the radar of growth capital providers such as VCs, at least outside the original industry cluster. As a result, promising startups tend to either seek acquisition by a multinational or relocate to superstar regions where growth capital is available. This can be seen in both Uppsala and Galway. One strategy that can help second-tier regions retain homegrown companies is to establish an exchange that allows small companies to go public relatively early. Examples of such venture exchanges include First North, which has launched in Denmark, Sweden, Iceland, and Finland, and the TSX Venture Exchange in Canada (Carpentier et al., 2010). The introduction of tax incentives or other monetary benefits for growth capital investors may also increase the attractiveness of investing in regional companies scaling up.

By retaining ventures with scaling potential, a second-tier region should aim to create an outstanding indigenous company that can serve as an inspiring example of local success. Such examples, as highlighted by interviewees from Galway and Uppsala, can inspire others to pursue entrepreneurial endeavors, creating a positive cycle of local entrepreneurial achievement. In addition, successful entrepreneurs sometimes become serial entrepreneurs or use their accumulated wealth to invest in other local startups, a process known as entrepreneurial recycling (Mason & Harrison, 2006).

As emphasized throughout this article, policymakers must continually adapt the policy framework and support mechanisms to ensure the sustainability and resilience of the EE. This will ensure that the needs of key stakeholders within the local EE are addressed to prevent the EE from entering

Growth	
Policy Approach	**Impact**
Encourage university-industry R&D collaboration via funding mechanisms	Attract and keep MNEs and indigenous ventures to/in the region, encourage spin-off activities
Ensure sufficient third-level funding	Ensure the availability of highly skilled human capital and local R&D capabilities
Facilitate research commercialisation via technology transfer offices at universities and IP regulation	Encourage spin-off venture creation by researchers
Ensure alignment between university research and teaching focus with local industry cluster	Ensure availability of highly skilled human capital for local economy and university-industry collaboration potential
Facilitate new venture creation via legal framework and infrastructure investments, and enable regional, national and international networking	Encourage spin-off venture creation by employees of large organisations and researchers at local universities
Develop and promote the capacity and the technological capabilities of emerging indigenous companies	Attract more FDI by incentivising foreign investors to further tap into the geographically concentrated local capabilities
Ensure sufficient funding for key actors within the local EE	Enable the continuous growth of the local EE
Promote the growing EE and its supportive infrastructure on a national and international level	Attract relevant actors, incl. nascent and early-stage entrepreneurs, to the region
Encourage collaboration among ecosystem actors, e.g. by creating attractive physical spaces that ensure density	Foster community development and a Pay-It-Forward type culture
Analyse the strengths and weaknesses of competing regional EEs	Ensure sustainable competitive advantage development relative to competing regions
Ensure the availability of relevant resources to current and nascent entrepreneurs and their awareness thereof	
Develop industrial, regional, science, and technology policy mechanisms that allow for technology and industry diversification	Ensure resilience of the growing EE, avoid technological lock-in, cluster hollowing-out and over-specialisation

Note: a dotted line indicates a policy approach already mentioned in a previous development stage that has been included again to emphasise its fluidity and relevance observed across development stages in flourishing second-tier EEs.

Figure 21: Policy Approaches for Second-Tier Entrepreneurial Ecosystem Growth. Source: own illustration.

Sustainment	
Policy Approach	**Impact**
Develop industrial, regional, science, and technology policy mechanisms that allow for technology and industry diversification	Ensure resilience of the growing EE and avoid technological lock-in, cluster hollowing-out and over-specialisation
Seek international networks and exchange ideas and experiences with other second-tier EEs	Generate ideas to avoid technological lock-in and diversify the local EE to ensure its resilience
Establish a small-cap venture exchange	Enable indigenous ventures to access growth capital and prevent them from MNE acquisition or relocating
Introduce tax incentives and/or other monetary benefits for venture capital investors and other growth capital providers	Attract venture capital investors and other growth capital providers to the region and thereby increase the availability of growth capital to local scale-ups
Use local success stories for promotional activities	Inspire the local population to start entrepreneurial ventures and attract entrepreneurs to the region
Ensure the wellbeing of local entrepreneurs by engaging in conversations and providing them with what they need	Foster entrepreneurial recycling and trigger a virtuous cycle of local entrepreneurial success
Adapt the policy framework and support mechanisms to cater for the needs of an increasingly diversified EE	Ensure the resilience of the local EE and prevent its decline

Note: a dotted line indicates a policy approach already mentioned in a previous development stage that has been included again to emphasise its fluidity and relevance observed across development stages in flourishing second-tier EEs.

Figure 22: Policy Approaches for Second-Tier Entrepreneurial Ecosystem Sustainment. Source: own illustration.

the decline phase. A list of policy approaches for sustaining an EE in second-tier European regions is presented in Figure 22.

The above policy approaches, their underlying mechanisms, and their respective effects were observed in existing second-tier EEs. However, as highlighted throughout the article, each EE is subject to a unique spatial context that must be considered when attempting to build, grow, and sustain a thriving ecosystem to support entrepreneurial activity. The next section summarizes the findings of this article and discusses the limitations of the results.

6. Conclusion and Limitations

Through a thorough literature review and the creation of a comparative case study of two regions, this article sought to derive dynamic policy implications for second-tier European regions seeking to effectively foster a local entrepreneurial ecosystem. While the importance of spatial differences was emphasized throughout, there are certain patterns that recur not only in the two second-tier European regions analyzed in this article, but also in the second-tier North American regions previously analyzed in the academic literature. First, it became clear that the establishment of thriving EEs in second-tier regions is possible despite polarization trends in innovation, but the emergence of an EE need not be intentional. Both Uppsala and Galway benefited from certain processes that later enabled the emergence of their respective EE. Thus, their creation was more accidental and a by-product of university-industry collaboration in Uppsala and attracting foreign direct investment in Galway. Both processes required government activity, but this varied between the regions. In Uppsala, the government was involved primarily through funding for higher education, while in Galway, the government created, among other things, a policy framework that provided incentives for multinational enterprises to locate in the region.

Second, the importance of R&D activities of local MNE subsidiaries and higher education institutions cannot be overstated. It was the transition from manufacturing to R&D by the subsidiaries of the multinationals in Galway, driven by the specialization of research at one of the local universities, that enabled the emergence of the EE. In Uppsala, on the other hand, it was the research strength of one of the local universities that attracted large companies to the region in the first place. R&D was thus a prerequisite for the birth of the EEs in Uppsala and Galway.

Third, another result revolves around the impact of innovation polarization on later stages of EE development. Both Uppsala and Galway suffer from a lack of access to growth capital. In both regions, this leads to increased acquisitions by multinationals and the relocation of scale-up ventures to national or international superstar regions. This finding seems particularly important as it underscores that location decisions by entrepreneurs have a direct impact on the growth potential of ventures. In addition, limited access to growth capital likely increases costs for scale-up ventures not operating in the region's focus industry if they attempt to find investors locally. This is due to a lack of qualified and experienced investors for the company's specific target industry. Such an increase in the cost of money results in either lower investment amounts or the loss of a larger portion of the venture itself, neither of which is beneficial to the entrepreneur. Therefore, the availability of growth capital should be an important criterion for entrepreneurs' location decisions, which means that the availability of growth capital should be a priority for regional decision makers, especially during the growth and sustainment of an EE.

The results of this article have certain limitations. By comparing two EEs, a "small n" method (Roundy, 2019) was chosen, which limits the degree of generalizability of the results. Given the different circumstances in the two regions, the results are subject to spatial context. However, as Eisenhardt and Graebner (2007) argue, this method nonetheless allows for an assessment of the presence and consistency of findings across cases, which provides some degree of generalizability. Moreover, both EEs revolve around subsectors of the life sciences. Focusing on this specific industry may limit the transferability of certain findings to other industry foci. However, by combining the insights gained from the comparative case study with a literature review on EEs in general and EEs in second-tier regions in particular, a more holistic picture was drawn from which policy implications for other second-tier European regions can be derived. Nonetheless, the findings may be sector-specific to some extent.

By exploring dynamic policy implications for second-tier European regions seeking to establish, grow, and sustain an EE, this article sought to contribute to research on the importance of the spatial context of EEs, as well as on the life cycle of EEs. The findings can help local policymakers develop policy frameworks and support mechanisms that enable an EE to emerge, grow, and sustain in a second-tier region. Given the link between entrepreneurship and economic development, the results of this article can contribute to the economic and social well-being of regions as well as their urban development plans.

References

Acs, Z. J., Åstebro, T., Audretsch, D., & Robinson, D. T. (2016). Public policy to promote entrepreneurship: a call to arms. *Small Business Economics*, *47*, 35–51.

Acs, Z. J., Braunerhjelm, P, Audretsch, D. B., & Carlsson, B. (2009). The knowledge spillover theory of entrepreneurship. *Small Business Economics*, *32*, 15–30.

Alston, M. (2004). 'You don't want to be a check-out chick all your life': The out-migration of young people from Australia's small rural towns. *Australian Journal of Social Issues*, *39*(3), 299–313.

Aparicio, S., Audretsch, D., & Urbano, D. (2020). Does entrepreneurship matter for inclusive growth? The role of social progress orientation. *Entrepreneurship Research Journal*, *11*(4), 20190308.

Atkinson, R., Muro, M., & Whiton, J. (2019). The case for growth centers. *The Brookings Institution*. Retrieved March 13, 2020, from https://www.brookings.edu/research/growth-centers-how-to-spread-tech-innovation-across-america

Audretsch, D., Belitski, M., & Desai, S. (2015). Entrepreneurship and economic development in cities. *The Annals of Regional Science*, *55*, 33–60.

Audretsch, D., Mason, C., Miles, M. P., & O'Connor, A. (2018). The dynamics of entrepreneurial ecosystems.

Audretsch, D., Thurik, R., Verheul, I., & Wennekers, S. (2002). Understanding entrepreneurship across countries and over time. In *Entrepreneurship: Determinants and policy in a European-US comparison* (pp. 1–10). Springer.

Auerswald, P. E. (2015). Enabling entrepreneurial ecosystems: Insights from ecology to inform effective entrepreneurship policy. *Kauffman Foundation Research Series on city, metro, and regional entrepreneurship*.

Bahrami, H., & Evans, S. (1995). Flexible re-cycling and high-technology entrepreneurship. *California Management Review*, *37*(3), 62–89.

Baumol, W. J. (1990). Entrepreneurship: Productive, unproductive, and destructive. *The Journal of Political Economy*, *98*(5), 893–921.

Biernacki, P., & Waldorf, D. (1981). Snowball sampling: Problems and techniques of chain referral sampling. *Sociological methods & research*, *10*(2), 141–163.

Boddy, C. R. (2016). Sample size for qualitative research. *Qualitative Market Research: An International Journal*, *19*(4), 426–432.

Borissenko, Y., & Boschma, R. (2016). A critical review of entrepreneurial ecosystems: towards a future research agenda. (1630).

Cardoso, R. V. (2018). Localising urbanisation trajectories: Comparing the emergence of second-tier urban regions in Europe (1890–2011). *Urban Research & Practice*, *11*(3), 223–246.

Carpentier, C., L'her, J.-F., & Suret, J.-M. (2010). Stock exchange markets for new ventures. *Journal of Business Venturing*, *25*(4), 403–422.

Cavallo, A., Ghezzi, A., & Balocco, R. (2019). Entrepreneurial ecosystem research: Present debates and future directions. *International Entrepreneurship and Management Journal*, *15*, 1291–1321.

City of Uppsala. (2015). Uppsala: Facts about business and enterprise. Retrieved July 13, 2020, from https://international.uppsala.se/globalassets/naringsliv/fakta/factsfigures_eng_webb_ny.pdf

Clancy, J., Ryan, P., & Giblin, M. (2013). High technology clusters, flagship MNE subsidiaries, and mothership ecosystems: Specialisation, adaption and connectivity in staving off cluster malaise'. *35th DRUID Celebration Conference 2013*.

Coad, A., Daunfeldt, S.-O., Hölzl, W., Johansson, D., & Nightingale, P. (2014). High-growth firms: introduction to the special section. *Industrial and Corporate Change*, *23*(1), 91–112.

Cohen, B. (2006). Sustainable valley entrepreneurial ecosystems. *Business strategy and the Environment*, *15*(1), 1–14.

Coughlan, C. (2017). Inside views of Ireland's digital technology revolution. Retrieved July 15, 2020, from https://www.irishexaminer.com/business/inside-view-of-irelands-digital-technology-revolution-441911.html

Dahlgren, H., & Valentin, F. (2007). Shaken not stirred: Multinational takeovers in Sweden's pharmaceutical industry. *Copenhagen Business School, Biotech Business Working Paper No. 04-2007*.

Dalle, J.-M., Den Besten, M., & Menon, C. (2017). Using Crunchbase for economic and managerial research. *OECD Science, Technology and Industry Working Papers 2017/08*.

Desrochers, P., & Hospers, G.-J. (2007). Cities and the Economic Development of Nations: An Essay on Jane Jacobs' Contribution to Economic Theory. *Canadian Journal of Regional Science*, *30*(1), 119–134.

Dijkstra, L., Garcilazo, E., & McCann, P. (2013). The economic performance of European cities and city regions: Myths and realities. *European Planning Studies*, *21*(3), 334–354.

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

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of management journal*, *50*(1), 25–32.

Eisenhardt, K. M., Graebner, M. E., & Sonenshein, S. (2016). Grand challenges and inductive methods: Rigor without rigor mortis. *Academy of Management Journal*, *59*(4), 1113–1123.

Eliasson, G., & Eliasson, Å. (2006). The Pharmacia story of entrepreneurship and as a creative technical university–an experiment in innovation, organizational break up and industrial renaissance. *Entrepreneurship & Regional Development*, *18*(5), 393–420.

European Commission. (2000). *The European observatory for SMEs: Sixth Report*. Luxembourg: Office for Official Publication of the European Communities.

European Commission. (2013). *Entrepreneurship 2020 action plan: Reigniting the entrepreneurial spirit in Europe*.

Evers, N., & Giblin, M. (2017). Born globals and the medical technology cluster in the west of Ireland. *European Born Globals Job Creation in Young Internationals Businesses, New York, Routledge*, 105–127.

Färnstrand Damsgaard, E., & Thursby, M. C. (2013). University entrepreneurship and professor privilege. *Industrial and Corporate Change*, *22*(1), 183–218.

Feld, B. (2020). *Startup communities: Building an entrepreneurial ecosystem in your city*. John Wiley & Sons.

Feldman, M. P. (2014). The character of innovative places: entrepreneurial strategy, economic development, and prosperity. *Small Business Economics*, *43*, 9–20.

Fischer, M., & Maggetti, M. (2017). Qualitative comparative analysis and the study of policy processes. *Journal of Comparative Policy Analysis: Research and Practice*, *19*(4), 345–361.

Florida, R. (2003). Entrepreneurship, creativity, and regional economic growth. *The emergence of entrepreneurship policy: Governance, start-ups and growth in the U.S. knowledge economy*, 39–58.

Florida, R. (2005a). *Cities and the creative class*. Routledge.

Florida, R. (2005b). The world is spiky: Globalization has changed the economic playing field, but hasn't leveled it. *The Atlantic Monthly*, *296*(3), 48.

Florida, R., Adler, P., & Mellander, C. (2017). The city as innovation machine. *Regional Studies*, *51*(1), 86–96.

Florida, R., & King, K. M. (2016). *Rise of the global startup city: the geography of venture capital investment in cities and metros across the globe*. Martin Prosperity Institute, University of Toronto.

Florida, R., & Mellander, C. (2016). Rise of the startup city: The changing geography of the venture capital financed innovation. *California Management Review*, *59*(1), 14–38.

Folta, T., Delmar, F., & Wennberg, K. (2010). Hybrid entrepreneurship. *Management Sciences*, *56*(2), 253–269.

Foray, D., & Goenaga, B. X. (2013). The goals of smart specialisation. S3 Policy Brief Series No. 01/2013.

Fusch, P. I., & Ness, L. R. (2015). Are we there yet? Data saturation in qualitative research. *The Qualitative Report*, *20(9)*, 1408–1416.

Fylan, F. (2005). Semi-structured interviewing. *A handbook of research methods for clinical and health psychology*, 65–77.

Giblin, M., & Ryan, P. (2012). Tight clusters or loose networks? The critical role of inward foreign direct investment in cluster creation. *Regional Studies*, *46*(2), 245–258.

Glaeser, E. L., Ponzetto, G. A., & Tobio, K. (2011). Cities, skills and regional change. *Regional Studies*, *48*(1), 7–43.

Halcomb, E. J., & Davidson, P. M. (2006). Is verbatim transcription of interview data always necessary? *Applied nursing research*, *19*(1), 38–42.

Henrekson, M., & Johansson, D. (2010). Gazelles as job creators: a survey and interpretation of the evidence. *Small Business Economics*, *35*, 227–244.

Hospers, G.-J. (2006). Jane Jacobs: her life and work. *European Planning Studies*, *14*(6), 723–732.

Isenberg, D. (2010). How to start an entrepreneurial revolution. *Harvard Business Review*, *88*(6), 40–50.

Isenberg, D. (2011a). Introducing the Entrepreneurship Ecosystem: Four Defining Characteristics. Retrieved April 22, 2020, from https://www.forbes.com/sites/danisenberg/2011/05/25/introducing-the-entrepreneurship-ecosystem-four-defining-characteristics/#7e8df8ce5fe8

Isenberg, D. (2011b). *The entrepreneurship ecosystem strategy as a new paradigm for economic policy: Principles for cultivating entrepreneurship*. Babson Park: Babson College.

Jacobs, J. (1969). *The New Economy of Cities*. Random House.

Kajornboon, A. B. (2005). Using interviews as research instruments. *E-journal for Research Teachers*, *2*(1), 1–9.

Kemeny, T., & Storper, M. (2020). Superstar cities and left-behind places: disruptive innovation, labor demand, and interregional inequal-

ity. *London School of Economics and Political Science, International Inequalities Institute Working Paper No. 41*.

Knight, G., & Cavusgil, S. (1996). The born global firm: A challenge to traditional internationalization theory'. In S. Cavusgil & T. Madsen (Eds.), *Advances in international marketing* (8th, pp. 11–26). Bingley: Emerald Group Publishing.

Kowal, S., & O'Connell, D. (2004). *The transcription of conversations* (U. Flick, E. Kardorff, & I. Steinke, Eds.). Thousand Oaks.

Kuckertz, A. (2019). Let's take the entrepreneurial ecosystem metaphor seriously! *Journal of Business Venturing Insights*, *11*, e00124.

Lee, N., & Rodríguez-Pose, A. (2021). Entrepreneurship and the fight against poverty in US cities. *Environment and Planning A: Economy and Space*, *53*(1), 31–52.

Lee, S. Y., Florida, R., & Ács, Z. J. (2004). Creativity and entrepreneurship: a regional analysis of new formation. *Regional Studies*, *38*(8), 879–891.

Leech, B. L. (2002). Asking questions: Techniques for semistructured interviews. *PS: Political Science & Politics*, *35*(4), 665–668.

Locke, E. A. (2007). The case for inductive theory building. *Journal of management*, *33*(6), 867–890.

Longhurst, R. (2003). Semi-structured interviews and focus groups. *Key methods in geography*, *3*(2), 143–156.

Mack, E., & Mayer, H. (2016). The evolutionary dynamics of entrepreneurial ecosystems. *Urban studies*, *53*(10), 2118–2133.

Maggioni, M. A. (2004). The rise and fall of industrial clusters: Technology and the life cycle of region. *IEB Working Paper 2004/06*.

Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample size in qualitative interview studies: guided by information power. *Qualitative Health Research*, *26*(13), 1753–1760.

Markusen, A. R., Lee, Y.-S., & DiGiovanna, S. (1999). Comprehending fast-growing regions. *Second tier cities: Rapid growth beyond the metropolis*, 3–19.

Mason, C., & Brown, R. (2014). Entrepreneurial ecosystems and growth oriented entrepreneurship. *Background paper prepared for the workshop organised by the OECD LEED programme and the Dutch Ministry of Economic Affairs on entrepreneurial ecosystems and growth oriented entrepreneurship (final version). The Hague: OECD*.

Mason, C., & Harrison, R. T. (2006). After the exit: Acquisitions, entrepreneurial recycling and regional economic development. *Regional studies*, *40*(1), 55–73.

Matheson, J. L. (2007). The Voice Transcription Technique: Use of Voice Recognition Software to Transcribe Digital Interview Data in Qualitative Research. *Qualitative Report*, *12*(4), 547–560.

Mayer, H. (2011). *Entrepreneurship and innovation in second tier regions*. Edward Elgar Publishing.

McLellan, E., MacQueen, K. M., & Neidig, J. L. (2003). Beyond the qualitative interview: Data preparation and transcription. *Field methods*, *15*(1), 63–84.

Moore, J. F. (1993). Predators and prey: a new ecology of competition. *Harvard Business Review*, *71*(3), 75–86.

Morgan, K. (2017). Nurturing novelty: Regional innovation policy in the age of smart specialisation. *Environment and Planning C: Politics and Space*, *35*(4), 569–583.

Morse, J. M. (1995). The significance of saturation. *Qualitative Health Research*, *5*(2), 147–149.

Motoyama, Y., Fetsch, E., Jackson, C., & Wiens, J. (2016). Little town, layered ecosystem: a case study of Chattanooga. *Layered Ecosystem: A Case Study of Chattanooga*.

Müller, S., & Korsgaard, S. (2018). Resources and bridging: the role of spatial context in rural entrepreneurship. *Entrepreneurship & Regional Development*, *30*(1-2), 224–255.

Muro, M., & Whiton, J. (2018). *Tech is (still) concentrating in the Bay Area: An update on America's winner-take-most economic phenomenon*. Retrieved March 13, 2020, from https://www.brookings.edu/blog/the-avenue/2018/12/17/tech-is-still-concentrating-in-the-bay-area-an-update-on-americas-winner-take-most-economic-phenomenon/

Nightingale, P., & Coad, A. (2014). Muppets and gazelles: political and methodological biases in entrepreneurship research. *Industrial and Corporate Change*, *23*(1), 113–143.

Nilsson, M., & Norell, K. (1996). *Entreeprenörsavknoppning från ett Storföretag – En studie av Pharmacia 1985–1996 (Entrepreneurial spin-off from a large company – A study of Pharmacia 1985–1996)*. Uppsala: Uppsvenska Handelskammaren.

OECD. (2013). Innovation-driven growth in regions: The role of smart specialisation (preliminary version. Retrieved May 26, 2020, from http://www.oecd.org/sti/inno/smart-specialisation.pdf

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research*, *42*, 533–544.

Parkinson, M., Meegan, R., Karecha, J., Evans, R., Jones, G., Sotarauta, M., Ruokolainen, O., Tosics, I., Gertheis, A., & Tönkő, A. (2012). SGPTD second tier cities and territorial development in Europe: Performance, policies and prospects. Final Report ESPON SGPTD project.

Patton, M. Q. (2005). Qualitative research. *Encyclopedia of statistics in behavioral science*, 1633–1636.

Porter, M. E. (1985). Competitive strategy: Creating and sustaining superior performance.

Pugh, R. E. (2014). 'Old wine in new bottles'? Smart Specialisation in Wales. *Regional Science*, *1*(1), 152–157.

Qian, H. (2018). Knowledge-based regional economic development: A synthetic review of knowledge spillovers, entrepreneurship, and entrepreneurial ecosystems. *Economic Development Quarterly*, *32*(2), 163–176.

Robinson, K. L., Dassie, W., & Christy, R. D. (2004). Entrepreneurship and small business development as a rural development strategy. *Proceedings of the Conference on Rural Community Economic Development: Experiences Fromthe 1890 Land-Grant Institutions*, *3-32*.

Rodríguez-Pose, A. (2018). The revenge of the places that don't matter (and what to do about it). *Cambridge Journal of Regions, Economy and Society*, *11*(1), 189–209.

Rosenthal, S. S., & Ross, A. (2010). Violent crime, entrepreneurship, and cities. *Journal of Urban Economics*, *67*(1), 135–149.

Roundy, P. T. (2017). "Small town" entrepreneurial ecosystems: Implications for developed and emerging economies. *Journal of Entrepreneurship in Emerging Economies*, *9*(3), 238–262.

Roundy, P. T. (2019). "It takes a village" to support entrepreneurship: Intersecting economic and community dynamics in small town entrepreneurial ecosystems. *International Entrepreneurship and Management Journal*, *15*(4), 1443–1475.

Roundy, P. T., Bradshaw, M., & Brockman, B. K. (2018). The emergence of entrepreneurial ecosystems: A complex adaptive systems approach. *Journal of Business Research*, *86*, 1–10.

Ryan, G. W., & Bernard, H. R. (2000). Data management and analysis methods. *Handbook of qualitative research*, 769–802.

Ryan, P., Buciuni, G., Giblin, M., & Andersson, U. (2020). Subsidiary upgrading and global value chain governance in the multinational enterprise. *Global Strategy Journal*, *10*(3), 496–519.

Ryan, P., & Giblin, M. (2012). High-tech Clusters, Innovation Capabilities and Technological Entrepreneurship: Evidence from Ireland. *The World Economy*, *35*(10), 1322–1339.

Ryan, P., Giblin, M., Buciuni, G., & Kogler, D. F. (2021). The role of MNEs in the genesis and growth of a resilient entrepreneurial ecosystem. *Entrepreneurship & Regional Development*, *33*(1-2), 36–53.

Sandström, A. (2014). *Global trends with local effects-The Swedish Life Science Industry 1998-2012*. VINNOVA-Swedish Governmental Agency for Innovation Systems.

Schmidt, C. (2004). The analysis of semi-structured interviews. *A companion to qualitative research*, *253*(41), 258.

Shwetzer, C., Maritz, A., & Nguyen, Q. (2019). Entrepreneurial ecosystems: a holistic and dynamic approach. *Journal of Industry-University Collaboration*, *1*(2), 79–95.

Spigel, B. (2017). The relational organization of entrepreneurial ecosystems. *Entrepreneurship Theory and Practice*, *41*(1), 49–72.

Spradley, J. (1979). *The ethnographic interview*. Holt, Rinehart; Winston.

Stake, R. E. (1995). *The art of case study research*. SAGE.

Stake, R. E. (2005). *Qualitative case studies.* SAGE.

Stam, E. (2015). Entrepreneurial ecosystems and regional policy: a sympathetic critique. *European Planning Studies*, *23*(9), 1759–1769.

STUNS Life Science. (2016). Life science in Uppsala. Retrieved July 13, 2020, from https://www.uppsalabio.com/facts-figures/uppsala/

Szerb, L., Acs, Z. J., Komlosi, E., & Ortega-Argilés, R. (2013). *REDI: The Regional Entrepreneurship and Development Index – Measuring regional entrepreneurship*. Brussels: European Commission, Directorate-General for Regional; Urban policy.

Szerb, L., Ács, Z. J., Komlósi, É., & Ortega-Argilés, R. (2015). Measuring entrepreneurial ecosystems: The regional entrepreneurship and development index (redi). *Discussion Paper No. CFE-2015-02. Henley Centre for Entrepreneurship*.

Szerb, L., Ortega-Argilés, R., Acs, Z. J., & Komlósi, É. (2020). Optimizing entrepreneurial development processes for smart specialization in the European Union. *Papers in Regional Science*, *99*(5), 1413–1457.

Teigland, R., Hallencreutz, D., & Lundequist, P. (2007). Uppsala BIO-The Life Science Initative: Experiences and Reflections on Starting a Regional Competitiveness Initiative. In *Small Business Clustering Technologies: Applications in Marketing, Management, IT and Economics* (pp. 198–223). IGI Global.

Teigland, R., Lindqvist, G., Malmberg, A., & Waxell, A. (2004). Investigating the Uppsala biotech cluster. *CIND Research Paper 2004:1. Uppsala universitet*.

Van Egeraat, C., & Jacobson, D. (2004). The rise and demise of the Irish and Scottish computer hardware industry. *European Planning Studies*, *12*(6), 809–834.

Vedula, S., & Kim, P. H. (2019). Gimme shelter or fade away: The impact of regional entrepreneurial ecosystem quality on venture survival. *Industrial and Corporate Change*, *28*(4), 827–854.

Wachsmuth, D. (2008). *Housing for immigrants in Ontario's medium sized cities*. Canadian Policy Research Networks Inc.; Social Housing Services Corporation.

Waxell, A., & Malmberg, A. (2007). What is global and what is local in knowledge-generating interaction? The case of the biotech cluster in Uppsala, Sweden. *Entrepreneurship and Regional Development*, *19*(2), 137–159.

Welch, M. (2011). The evolution of the employee engagement concept: communication implications. *Corporate Communications: An International Journal*, *16*(4), 328–346.

Wellard, S., & McKenna, L. (2001). Turning tapes into text: Issues surrounding the transcription of interviews. *Contemporary Nurse*, *11*(2-3), 180–186.

Welter, F., Trettin, L., & Neumann, U. (2008). Fostering entrepreneurship in distressed urban neighbourhoods. *International Entrepreneurship and Management Journal*, *4*, 109–128.

Xu, Z., & Dobson, S. (2019). Challenges of building entrepreneurial ecosystems in peripheral places. *Journal of Entrepreneurship and Public Policy*, *8*(3), 408–430.

Junior Management Science 9(1) (2024) 1241-1268

Junior Management Science

www.jums.academy
ISSN: 2942-1861

Cost Allocation in Vehicle Routing Problems with Time Windows

Federico Arroyo

Technical University of Munich

Abstract

The estimation of costs allocated to each customer when serving them in a collaborative logistic operation is a complex problem whose solution is computationally very expensive. In this work the case of central horizontal collaboration for vehicle routing problems with time windows and a central depot is studied. An approximation to the Shapley value method via structured random sampling is used to calculate the cost associated with customers in Solomon instances. Such costs are regressed to a linear model with a set of defined features. The results show that cost can be predicted with considerable accuracy with few features. Moreover, the extent to which vehicles' capacity, customers' demand and distance, the degree of customer clustering and time window horizons affect cost and potential savings from carriers in collaboration is assessed. Additionally, individual regression models of different set of instances show how various pricing strategies for customers can be fitted to their classification when grouping them.

Keywords: collaborative vehicle routing; cost allocation; Shapley value method; structured random sampling; time windows

1. Introduction

1.1. Context and motivation

With the world's population continuing to grow and an ever-expanding international commerce, logistics networks become increasingly complex. The supply chain, as a means of delivering products from manufacturers to customers, comprises the interactions between enterprises involving a multi-organizational network adding value to the flow of information, logistics and capital (Leng et al., 2018; Tang et al., 2022). It is within this operation that any potential for improvement brings substantial value.

Further, the mass adoption of the internet, in conjunction with the development of new technologies, has made online shopping a very comfortable means for consumers to acquire their products, favoring it in many cases over brick-and-mortar shops. This change of behavior towards ordering products online has meant rapid growth for e-commerce which has been further exacerbated by the COVID-19 pandemic, feeding the need for more retailers online and the growth in deliveries (Alfonso et al., 2021; Jílková & Králová, 2021). The final leg of the logistics problem, known as the last mile, is where retailers and logistic carriers mostly find their costs increasing due to the inefficiencies of the industry (Iannaccone et al., 2021). This operation of delivering a parcel from the last hub to the customer accounts for 28 % of the total logistic cost (Ranieri et al., 2018). Moreover, the effect of increased urban parcel demands in the context of urbanization can be very harmful to cities. With 70 % of the world's population expected to live in cities by 2050, the impact on the environment, safety and health from pollution and congestion becomes even more important (Bretzke, 2013). Consequently, public institutions and governments are beginning to regulate carriers' operation more intensely, pushing for in-

I would like to express my gratitude to Prof. Dr. Stefan Minner for affording me the invaluable opportunity to compose my master thesis under his supervision and mentorship. His continuous provision of insightful suggestions, innovative ideas to enhance my thesis, and unwavering support during the writing process have been truly instrumental. Furthermore, my appreciation extends to the esteemed Chair of Logistics and Supply Chain Management of the Technical University of Munich where I was given full assistance and support along my studies and in preparation for my thesis. Lastly, I extend my thanks to family and friends, whose steadfast encouragement and caring have been a driving force behind my endeavors, fostering both my personal and professional growth.

DOI: https://doi.org/10.5282/jums/v9i1pp1241-1268

creased sustainability and efficiency in their operations (Hu et al., 2019). With regards to the timing of deliveries, first-time delivery failures have been reported up to 60 %, resulting in a significant cost for carriers (Song et al., 2009). Additionally, same-day deliveries and the expectations coming from customers of receiving their parcels fitting their schedules compel agents to work with tight time horizons.

Within this context, companies are either forced or need to collaborate with each other to remain competitive and streamline their costs. Horizontal collaboration appears as an effective solution for carriers to increase their operational efficiency (Gansterer & Hartl, 2018b). For this reason, it is increasingly important for public agents in the last years (Cruijssen & Cruijssen, 2020). In a sharing economy, where companies are using common resources and capacities in urban areas to provide new services, new business models emerge and traditional ones need to innovate in order to continue doing business (Dahle et al., 2019).

1.2. Problem statement

In recent years, the area of collaborative vehicle routing has gained great popularity in transportation and logistics research (Gansterer & Hartl, 2020). Shared transportation resources define the base concept where different partners join in a common operation. In the realm of collaborative vehicle routing, we focus on those problems with capacity and time constraints, later explained in Section 3.1. To cope with the growth of deliveries and the ever-tightening constraints of time and capacity, collaborative environments allow for a great level of optimization within logistic operations (Karaenke et al., 2019; Vanovermeire et al., 2014). In this regard, there is a strong incentive for carriers to participate in a joint coalition given the cost reductions they can achieve (Muñoz-Villamizar et al., 2015).

Moreover, through the optimized use of resources, companies can benefit from providing better service, due to their increased flexibility and better times of delivery (Cruijssen, Dullaert, & Fleuren, 2007). Additionally, there is a need for them to decrease their environmental impact. In the USA, for example, it has been found that logistics vehicles can travel totally empty in up to 15% to 20% of the distance they cover (Ferrell et al., 2020). Companies, therefore have the motivation to increase their sustainability and their reputation.

Within the paradigm of companies cooperating and sharing resources, we focus on horizontal collaboration where carriers belong to the same level of the supply chain. Vertical collaboration, on the other hand, involves other levels in it. A common understanding, represented through a central system or platform, is needed for agents to interact and plan together horizontally, with some cities already introducing such systems (Schmelzer et al., 2016). The way collaboration structures are defined is crucial to allow for horizontal collaborations to prosper due to competitors being reticent to share their information and resources for lack of trust (Bratton et al., 2000). Transparency and confidentiality are key for competitors not to fear their strategic advantage is threatened. Consequently, having a fair distribution of costs arisen and profits obtained is of greater importance for the stakeholders involved (Cruijssen, Cools, & Dullaert, 2007). In this regard, our work is centered in cost allocation in collaborative vehicle routing problems, where the total cost of a global operation needs to be divided between the individual carriers. The objective is to assign costs to each of the customers that each partner brings into the coalition in a simple but fair way. The cost allocation method must assure that no partner perceives their allocated cost as being too large, for they risk the stability of the cooperation. We explore this in Section 3.2. The numerical complexity of the problem calls for the search for simpler mechanisms which is at the center of our work.

1.3. Research question and thesis outline

Given the scope presented, we aim at answering the following question:

> *Can cost allocation mechanisms in vehicle routing problems with time windows be approximated through simple methods?*

With this, we determine the accuracy of a simple cost-allocating linear model where certain features are used to address the variables of the problem. Additionally, by using different instances with different degrees of clustering and time window horizons, we analyze the impact these dimensions have on the cost-allocating process and derive conclusions for different pricing strategies. Concerning the allocation mechanism itself, we determine the degree to which the allocating results are advantageous for all carriers and whether or not all customers yield savings when shared with others.

This work begins by reviewing the available literature on the subject and summarizing basic definitions in Section 2. The methodology to tackle the problem in question is defined in Section 3 where the modelling and solving of vehicle routing problems, cost-allocating mechanisms and regression methods are presented. Later, in Section 4, the analyzed instances and cost function features are introduced. Subsequently, the results obtained are analyzed in Section 5. This comprises an overview of the costs allocated to each customer for the different instances, followed by the model fit and feature selection and the study of collaborative gains. To finalize, conclusions and recommendations for future research directions are pointed out in Section 6.

2. Literature review

In the context of our investigation, we begin our work by thoroughly reviewing the current state of academic research concerning cost allocation in collaborative multi-carrier vehicle routing problems with time windows. To better digest the subject, we focus on the two main bodies of research that our study focuses on. On the one hand, we start by evaluating the vehicle routing problem and its variations; focusing on horizontal collaboration. Further, we describe the related

work in the area of cost allocation methods for collaborating logistics. We aim to present the known sources on which we base our work, finding potential unexplored areas in research that we intend to develop.

2.1. The vehicle routing problem

The Vehicle Routing Problem (VRP), introduced by Dantzig and Ramser (1959) and Clarke and Wright (1964), still remains one of the most widely studied problems in operations research (Kritikos & Ioannou, 2010). It consists of a family of problems in which a set of customers with known demands and geographical location have to be served by a fleet of vehicles based at one (or many) depot(s) through a network of paths between them. The task is to determine the optimal set of routes for the vehicles which achieve the objective and satisfy the constraints imposed on the problem (Irnich et al., 2014).

Multiple variations and extensions to the VRP have been introduced which are mainly classified based on their objectives and constraints (Laporte, 2009). Most commonly, a capacity restriction is imposed on the vehicles, with each location having a specified demand quantity. This problem, known as the Capacitated VRP (CVRP), imposes that each customer needs to be served only once and the capacity of vehicles is not surpassed by the total amount of customer demands served by the vehicle in its route (Goel & Gruhn, 2008; Prins, 2004). Additionally, the CVRP with Time Windows (CVRPTW) introduces the time dimension and the constraint of customers being served in a time interval, also known as Time Window (Kallehauge et al., 2005). The addition of several depots with multiple pickup and delivery points (Berbeglia et al., 2007) as the discrimination between full truckload and less than truckload (Gansterer & Hartl, 2018b) are also well-known alternatives that we will not explore in our work.

When several carriers are present, each of them having to solve a VRPs by their own, potential benefits have been found when collaborating in a joint logistic operation, originally reviewed in (Cruijssen, Dullaert, & Fleuren, 2007; Krajewska & Kopfer, 2006).

In their survey, Gansterer and Hartl (2018b) show that problems in vehicle routing in a collaborative scenario can be addressed from a central or decentral planning perspective. Central plans require the intervention of one decision-maker who is fully informed. The information provided for each carrier can consist of cost structures, capacities and information on existing customers (Gansterer & Hartl, 2018b). An example of this in digitalized times would be an online platform/database where carriers' requests are shared and solved in a logistics problem. In this situation, the problem is reduced to a holistic standard optimization problem where all resources are bundled and the collaborative aspect is overtaken by the full disclosure of information.

In decentralized collaborations, carriers are able to share their capacities to a certain extent by revealing a limited amount of information. In said setting partners may deal individually with each other or channel their requests through a central authority. Studies on this area are presented by (Berger & Bierwirth, 2010; Dai & Chen, 2011; Gansterer & Hartl, 2018a; Gansterer et al., 2019; Krajewska & Kopfer, 2006; Özener et al., 2011), covering non-auction and auction-based mechanisms for the selection and exchange of requests. As reviewed by Gansterer and Hartl (2018b), when it comes to real-world cases the potential collaboration gains have been calculated up to 20%-30% (Chinh et al., 2016; Cruijssen, Bräysy, et al., 2007; Ergun et al., 2007; Muñoz-Villamizar et al., 2015). In the case of each customer requiring service by two or more carriers Fernández et al. (2018) obtained cost savings of $6\% - 25\%$. Next to the economical aspect, the nature of collaboration between carriers signifies a reduced negative impact on the environment. These aspects are found in the research by Ballot and Fontane (2010), Muñoz-Villamizar et al. (2015), Pérez-Bernabeu et al. (2015), and Schulte et al. (2017) which study the reduction of routes and vehicles leading to a decrease in CO2 emissions and congestion in cities.

From the numerical solving perspective, collaborative VRPs encompass a great complexity which requires the use of heuristic and metaheuristic approaches. With regard to the VRP, the solving mechanisms are found in Cordeau et al. (1997, 2007), Laporte (1992), and Tarantilis (2005). Collaborative VRPs solving algorithms are discussed by Defryn, Sörensen, and Cornelissens (2016), Pérez-Bernabeu et al. (2015), and Sanchez et al. (2016).

2.2. Cost allocation methods

In the scenario of collaboration, the profits gained by the coalition and the costs assumed have to be shared among the participants. A transparent and fair mechanism to split these values is necessary in order to assure the collaborative relationships between the partners and their long-term participation (Defryn, Sörensen, & Cornelissens, 2016).

In the context of the VRP and its variations, the cost allocation problem consists of splitting the cost of the route/s between the customers served. This was first addressed by Göthe-Lundgren et al. (1996). Some other practical applications were developed by Engevall et al. (2004), Krajewska and Kopfer (2006), and Krajewska et al. (2008). For CVRPTW, cost allocation was first studied by Cruijssen et al. (2010) and later analyzed by Dahl and Derigs (2011) in a dynamic setting.

In their survey, Guajardo and Rönnqvist (2016) show that in most cases in literature cooperative game theory is applied for the sharing of costs or profits. The authors categorize more than 40 methods either as traditional or ad-hoc. The first one refers to methods arising from previous work on cooperative game theory whereas ad-hoc are methods which result out of the particular conditions of the problem. Nonetheless, they conclude that a big majority of the methods found in literature fall into the following:

- the Shapley value (Shapley, 1953), generally the most applied method which we explain later;

- proportional methods, where the cost is divided proportionally to a variable to each carrier (e.g. carrier j is allocated cost α_j). The share of cost assigned to each player can follow different criteria (e.g. shared equally among players, according to their demand quantities or standalone costs);
- the nucleolus method (Schmeidler, 1969), which looks for an individually rational distribution in which the maximum dissatisfaction is minimized.

Table 1 extracted from the literature survey by Guajardo and Rönnqvist (2016) summarizes the methods that are of most relevance for our work. From the traditional methods, the Shapley value is the one we focus on given its broad use. The methods consolidated under the category "ad-hoc" respond to those following definitions which concern the particular context where the research has been motivated, including modifications of traditional methods.

In their work Shi et al. (2020) analyse the problem of sharing profits in collaborative vehicle routing problems with multiple depots. A similar approach is what we apply for splitting costs in problems with rather one depot and the added constraint of time-windows. Akkerman and Mes (2022) analyze how customer selection in vehicle routing problems could be approximated by the distance. We expand this area of study by addressing which other features, and to which extent, could be used to approximate cost being allocated to customers. To our best knowledge, there is no academic work where the cost allocation for collaborative environments of CVRPTW is addressed through approximation methods using linear regression. This is what we wish to explore with our work.

3. Methodology

As the focus of this work is to analyze the applicability of direct rules and approximate a formula for the cost allocation of clients in a collaborative vehicle routing problem, we will develop our work by evaluating different CVRPTW Solomon data instances (Solomon, 1987) and assessing how the measures designed fit the outcome obtained. Firstly, we model the logistics problem as a vehicle routing problem with the added constraints of vehicle capacity and time windows and build our optimization model. In order to streamline the computation of our results, a metaheuristic solver is introduced. Further, we assess different methods for the allocation of costs in collaborative games. In particular, the full calculation of the Shapley value is initially adopted and later replaced with a structured random sample-based method. Finally, our model is applied to existing problem instances so as to derive the results which are later analyzed.

3.1. Modelling and solving of vehicle routing problems

3.1.1. Capacitated Vehicle Routing Problem (CVRP)

For the modelling of our problem, we start with a simpler case in the CVRP. In this regard, the two-index flow formulation by Toth and Vigo (2014) is used. The CVRP aims at finding the set of routes that allow all customers' demands in a network to be serviced by a fleet of vehicles starting from a depot with the minimum cost. Further, customers' demands and geographical locations are deterministic with the latter deriving in the network of possible arcs available between nodes. Vehicles have a limit to the capacity that they can carry. In this regard, the notation is defined in the following.

Let $V = 0,\dots,n$ be the vertex set, where vertices $i = 1,\dots,n$ correspond to the customers and vertex 0 corresponds to the depot. A is then the arc set which completes the graph $G = (V,A)$, with each arc having a cost c_{ij} associated with it: $(i,j) \in A$. Furthermore, each customer has a demand d_i to be fulfilled by a set of K available vehicles (K being not smaller than the minimum amount of vehicles needed to serve all customers), which are indistinguishable with the same capacity C measured in the same units as the demand. For a given circuit, the cumulative demand of all the customers served by it must not exceed the vehicle capacity. Finally, given a subset $S \subseteq V\backslash\{0\}$, we denote by $r(S)$ the minimum number of vehicles needed to serve all customers in S.

In all the instances we use throughout our work, vertices are defined by points with specific coordinates. Further, the cost c_{ij} linked to each arc $(i,j) \in A$, is calculated as the Euclidean distance between the two points corresponding to the vertices i and j. This results in a cost matrix which is symmetric and satisfies the triangle inequality:

$$c_{ik} + c_{kj} \geq c_{ij} \quad \forall i,j,k \in V$$

which signifies that any deviation from the direct link between two vertices results in a bigger distance, therefore bigger cost. In the remaining of this work, distance will be used as a measure of cost and vice-versa.

The model uses binary variables x_{ij} which take the value of 1 if the arc (i,j) is traversed in the solution and 0 otherwise. With the notation described, the problem formulation is as follows:

$$\min \sum_{i\in V}\sum_{j\in V} c_{ij}x_{ij} \tag{1a}$$

$$\text{s.t.:} \sum_{i\in V} x_{ij} = 1 \quad \forall j \in V\backslash\{0\}, \tag{1b}$$

$$\sum_{j\in V} x_{ij} = 1 \quad \forall i \in V\backslash\{0\}, \tag{1c}$$

$$\sum_{i\in V} x_{i0} = K, \tag{1d}$$

$$\sum_{j\in V} x_{0j} = K, \tag{1e}$$

$$\sum_{i\notin S}\sum_{j\in S} x_{ij} \geq r(S) \quad \forall S \subseteq V\backslash\{0\}, S \neq \emptyset, \tag{1f}$$

$$x_{ij} \in \{0,1\} \quad \forall i,j \in V. \tag{1g}$$

The objective function in equation 1a aims at minimizing the sum of costs deriving from traversing the routes. Constraints 1b and 1c ensure that each customer is visited only

Table 1: Cost allocation methods in a collaborative environment found in academic literature (Guajardo & Rönnqvist, 2016).

Method	No. of articles	References
Ad hoc	31	(Agarwal & Ergun, 2010; Anily & Haviv, 2007; Audy et al., 2011, 2012; Caprara & Letchford, 2010; Cheng et al., 2013; Dahl & Derigs, 2011; Dai & Chen, 2012, 2015; Derks & Kuipers, 1997; Engevall et al., 1998, 2004; Estévez-Fernández et al., 2009; Faigle et al., 1998; Fiestras-Janeiro et al., 2013; Fiestras-Janeiro et al., 2012; Flisberg et al., 2015; Frisk et al., 2010; Guajardo & Rönnqvist, 2015; Hamers et al., 1999; Hezarkhani et al., 2016; Liu et al., 2010; Özener, 2014; Özener & Ergun, 2008; Özener et al., 2013; Potters et al., 1992; Sun et al., 2015; Toriello & Uhan, 2013; Vanovermeire & Sörensen, 2014b; Yang et al., 2016; Yilmaz & Savasaneril, 2012)
Shapley	23	(Agarwal & Ergun, 2010; Cheng et al., 2013; Cruijssen et al., 2010; Dror, 1990; Engevall et al., 1998, 2004; Fang & Cho, 2014; Fiestras-Janeiro et al., 2012; Frisk et al., 2010; Guajardo et al., 2016; Hezarkhani et al., 2016; Krajewska et al., 2008; Liu et al., 2010; Lozano et al., 2013; Massol & Tchung-Ming, 2010; Özener et al., 2013; Sun et al., 2015; Vanovermeire & Sörensen, 2014a; Vanovermeire et al., 2014; Wang et al., 2015; Wong et al., 2007; Yengin, 2012; Zakharov & Shchegryaev, 2015)
Proportional	18	(Audy et al., 2012; Berger & Bierwirth, 2010; Dror, 1990; Engevall et al., 2004; Fishburn & Pollak, 1983; Flisberg et al., 2015; Frisk et al., 2010; Hezarkhani et al., 2016; Krajewska & Kopfer, 2006; Lehoux et al., 2011; Liu et al., 2010; Massol & Tchung-Ming, 2010; Nguyen et al., 2014; Özener, 2014; Özener & Ergun, 2008; Özener et al., 2013; Sun et al., 2015; Wong et al., 2007)

once allowing for exactly one arc to enter and leave each vertex associated with them. Consequently, constraints 1d and 1e impose analog requirements to the depot. Note that with this notation the CVRP aims at finding a collection of exactly K circuits to serve all customers, where each circuit starts and ends at the depot. Constraints 1f ensure connectivity of the paths in the solution while also imposing the vehicle capacity requirements. The constraints remain valid also if $r(S)$ is replaced by a trivial lower bound defined by $\lceil d(S)/C \rceil$ (Cornuejols & Harche, 1993) with $d(S) = \sum_{i \in S} d_i$ denoting the total demand of the set $S \subseteq V$. Lastly, constraint 1g represents the binary constraint.

The family of constraints 1f has a cardinality which grows exponentially with n (Toth & Vigo, 2014). A family of equivalent constraints with polynomial cardinality can be used by utilizing the sub-tour elimination constraint of the Travelling Salesman Problem (TSP) applied to the CVRP (Christofides, 1979; Desrochers & Laporte, 1991; Miller et al., 1960):

$$u_i - u_j + C x_{ij} \leq C - d_j \quad \forall i, j \in V \backslash \{0\}, i \neq j \tag{2a}$$

$$\text{such that } d_i + d_j \leq C,$$

$$d_i \leq u_i \leq C \forall i \in V \backslash \{0\}, \tag{2b}$$

where the variable $u_i, i \in V \backslash \{0\}$ is added to the model, accounting for the load of the vehicle after visiting customer i. With this definition, when $x_{ij} = 0$, constraint 2a is not binding, while when $x_{ij} = 1$ then $u_j \geq u_i + d_j$ imposing the capacity and connectivity requirements and eliminating isolated sub-tours in the process (Toth & Vigo, 2014).

3.1.2. CVRP with Time Windows (CVRPTW)

The CVRP with Time Windows extends the CVRP further restricting the problem by adding a time dimension. Each customer is associated with a time interval in which it needs to be served, called a time window. Further, the travel time, t_{ij}, for each arc $(i, j) \in A$ is known and normally taken as the same as the distance or cost of the arc c_{ij}. Additionally, service times s_i are given for each customer i and the time instant in which the vehicles leave the depot is also known and normally assumed to be at 0. As a result, the service at each customer i has to be initiated during their time window and it lasts the service time associated with it s_i. Furthermore, whenever a vehicle arrives earlier than the start of a customer's time window, it is allowed to wait until it begins and then commence service.

For the purpose of this work and to summarize it's main characteristics, the classical CVRPTW consists of defining exactly one circuit per vehicle (of the total K vehicles available) minimizing the cost with:

- each vehicle starting and finishing at the depot,
- customers being serviced exactly once,
- the capacity of the vehicles C not exceeding the sum of the demands of the customers serviced by it,
- service starting within the time window $[a_i, b_i]$ of each customer and the vehicle remaining a time equal to s_i at location i.

The model used in our work for the CVRPTW follows the formulation by Toth and Vigo (2014), which is here described. It is based on the three-index vehicle flow model which allows for the modelling of more constrained versions of the vehicle routing problem, due to being more flexible in integrating additional dimensions.

Firstly, a network $G = (V,A)$ is defined by the set of vertices of customers and the depot, which in this case is represented by the two nodes 0 and $n+1$. The set of customers N is then defined as all vertices but vertices 0 and $n+1$, namely $N = V\backslash\{0, n+1\}$. Further, for a route to be feasible then it must start from node 0 and end at node $n+1$. Additionally, a time window $[a_i, b_i]$ is associated with each customer i. Nodes 0 and $n+1$ are assigned E and L as their earliest possible departure and the latest possible arrival at the depot respectively; that is, $[a_0, b_0] = [a_{n+1}, b_{n+1}] = [E, L]$. Moreover, and in accordance with the problem description, zero demands and service times are defined for these two nodes, namely $d_0 = d_{n+1} = s_0 = s_{n+1} = 0$.

Let $\Delta^+(i)$ denote the set of vertices j that are connected to the vertex i directly, such that arc $(i,j) \in A$. In the same manner, $\Delta^-(i)$ denote the vertices from which i is directly reached, e.g. the set of vertices j such that arc $(j,i) \in A$.

For each arc $(i,j) \in A$ and vehicle $k \in K$, a variable x_{ijk} is introduced which takes the value of 1 if arc (i,j) is used by vehicle k and 0 otherwise. Further, a time variable w_{ik} is added for all nodes $i \in V$ and vehicles $k \in K$ which designates the time of the start of service at node i by vehicle k.

With the described notation, the formulation as of Toth and Vigo (2014) is as follows:

$$\min \sum_{k\in K} \sum_{(i,j)\in A} c_{ij} x_{ijk} \tag{3a}$$

$$\text{s.t.:} \sum_{k\in K} \sum_{j\in\Delta^+(i)} x_{ijk} = 1 \quad \forall i \in N, \tag{3b}$$

$$\sum_{j\in\Delta^+(0)} x_{0jk} = 1 \quad \forall k \in K, \tag{3c}$$

$$\sum_{i\in\Delta^-(j)} x_{ijk} - \sum_{i\in\Delta^+(j)} x_{jik} = 0 \quad \forall k \in K, j \in N, \tag{3d}$$

$$\sum_{i\in\Delta^-(n+1)} x_{i,n+1,k} = 1 \quad \forall k \in K, \tag{3e}$$

$$x_{ijk}\left(w_{ik} + s_i + t_{ij} - w_{jk}\right) \leq 0 \quad \forall k \in K, (i,j) \in A, \tag{3f}$$

$$a_i \sum_{j\in\Delta^+(i)} x_{ijk} \leq w_{ik} \leq b_i \sum_{j\in\Delta^+(i)} x_{ijk} \quad \forall k \in K, i \in N, \tag{3g}$$

$$E \leq w_{ik} \leq L \quad \forall k \in K, i \in \{0, n+1\}, \tag{3h}$$

$$\sum_{i\in N} d_i \sum_{j\in\Delta^+(i)} x_{ijk} \leq C \quad \forall k \in K, \tag{3i}$$

$$x_{ijk} \in \{0,1\} \quad \forall k \in K, (i,j) \in A. \tag{3j}$$

The objective function calculates the total cost incurred which we aim to minimize. Constraints 3b ensure that each customer is assigned to exactly one route. Further, constraints 3c to 3e designate the flow on the path to be followed by vehicle k. After, constraints 3f to 3h ensure the time schedule is feasible while 3i guarantees the capacity constraint. Finally, constraints 3j ensure the binary conditions of the variables x_{ijk}.

3.1.3. Metaheuristic solver

The CVRP, as described in Section 3.1.1, is a generalization of the TSP (this being a CVRP with one vehicle visiting all vertices in a cost-minimizing circuit with $C \geq d(V)$ and $K = 1$). This problem and its extensions with Time Windows (CVRPTW) is known to be strongly NP-hard (Toth & Vigo, 2014). For this reason, it would be very expensive computationally to solve each problem to optimality. As a consequence, a solver algorithm using metaheuristics is implemented.

VROOM is an open-source optimization engine specifically tailored for VRPs which provides very good solutions in little time (VROOM Project, 2022). As an example of it performance, Table 2 shows the results obtained when solving Solomon 56 instances of 100 customers which are used for benchmarking. By comparing the cumulated traveled time, it is observed that the maximum gap difference with the best-known solution is 3.89%, all with an average computing time of 359 milliseconds.

The mentioned algorithm utilizes different heuristics depending on the problem to find an initial solution. In the case of the CVRPTW, it uses modified versions of Solomon insertion heuristics. To improve the solution, a local search procedure consisting of 14 different operators is then performed to check for better feasible neighboring solutions (Bräysy & Gendreau, 2005). In their work comparing tools for solving CVRPTW, Puka et al. (2019) find that VROOM in comparison with other open-source solvers is one of the most balanced in terms of solution quality and execution time. For all the aforementioned reasons, it is used in our research.

3.2. Cost allocation in VRPs in a collaborative environment

In this section, the vehicle routing problems previously defined are introduced in a collaborative environment in which several companies, each with a set of customers to serve, form a coalition that aims at satisfying the demand of all customers in a joint operation. In this setup, all customers and their demands are joined into a holistic logistic problem that is indifferent to the source of each customer. Operating as a joint venture, it has been shown that a total reduced cost is achieved while serving all customers. This is produced by the increased efficiency in the joint logistic operation due to customers with similar locations or service time windows being visited by the same vehicle. The cost that companies pay for such an optimization coalition is having to share their customer information which may be of high value for their business. We will consider a set of partners which form a grand coalition sharing their resources (vehicles) and demands. An example of such a collaborative CVRP is displayed in Figure 1 where 3 partners join in a coalition to serve all of their customers resulting in a completely different total logistic operation.

When working as a coalition of two or more partners, the problem arises of distributing the costs (or the profits) of serving the customers as a whole. In this regard, a mechanism is needed to allow for each customer that is brought

Table 2: Comparison to best known solutions (BKS) targeting cumulated travel time (CTT). Cumulated number of vehicles (CNV) is only reported for the record. (VROOM Project, 2022)

Class	C1	C2	R1	R2	RC1	RC2	Total
BKS CNV	10.0	3.0	13.25	5.36	12.88	6.25	485
BKS CTT	828.38	589.86	1175.75	878.41	1340.02	1004.21	54699
VROOM CNV	10.0	3.0	13.25	4.27	13.25	5.25	468
VROOM CTT	828.38	589.86	1191.99	912.60	1356.76	1030.75	55,616
CTT gap	+0.00%	+0.00%	+1.38%	+3.89%	+1.25%	+2.64%	+1.68

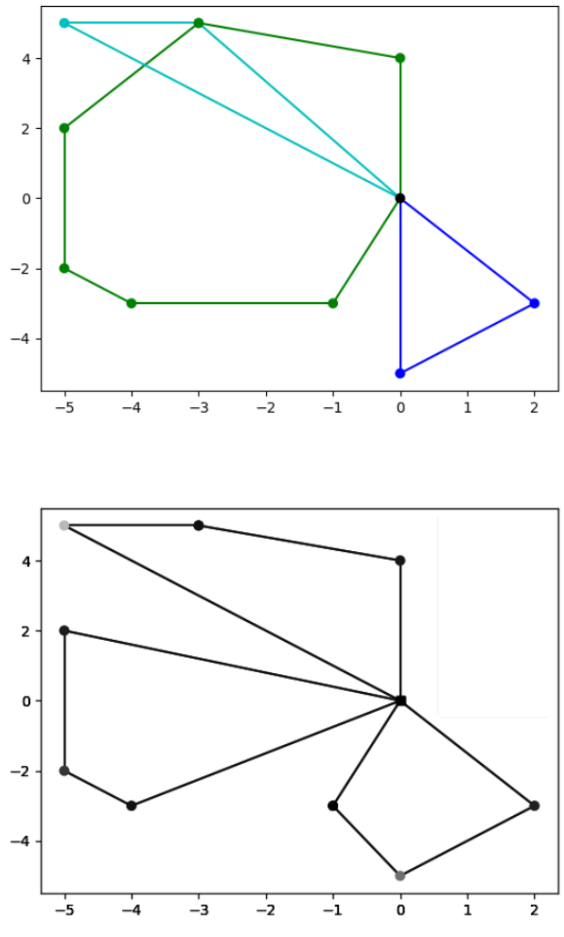


Figure 1: Example of a collaborative VRP where 3 partners (top green, cyan, blue) work jointly on a logistic operation.

in the coalition by a partner to be correctly accounted for the marginal cost it produced to the total operation. Despite no method being considered as a global best practice, Defryn, Vanovermeire, and Sörensen (2016) argue that a cost allocation method based on the incentives of each partner should be selected by the coalition. This results in the group behaving in the best interest of the whole, e.g.: a cost allocation based on volume will derive in benefit for those who transport the highest volumes therefore steering the coalition into increasing the total volume displaced. Nonetheless, there is a general consensus on the fact that incentive criteria results in a fair distribution of costs in all cases (Defryn, Sörensen, & Cornelissens, 2016). We follow with the introduction of the concepts behind cooperative game theory and an axiom-based cost allocation method: the Shapley value.

3.2.1. Cooperative game theory: Core

In game theory when evaluating a game, one is interested in knowing the value each player adds to a coalition when a group of players cooperates and obtains a certain overall gain from that cooperation. Some players' contributions may be larger or they may have a bigger bargaining power, even so to the extent of threatening to destroy the entire surplus. Therefore, the goal of this assessment would be to determine how important is each participant to the overall coalition, and what would therefore be their expected fair payoff. An analogous formulation applies to the case of determining the

cost of each customer added to a collaborative logistic problem as the only adjustment is that in the VRP each added player (customer) to the coalition will add a cost for the total operation instead of a surplus.

Let us introduce some cooperative game theory concepts as described by Guajardo and Rönnqvist (2015). For a cooperative game in which the "grand coalition" of all players in the set $N = \{1, \ldots, n\}$ participate, there is a characteristic function v which assigns a total cost for each coalition $S \in P$ with P comprising the set of all possible subsets of N. In our case, this cost is calculated by solving the corresponding optimization problem of the CVRPTW for the coalition S of customers (players). A cost allocation vector $z = (z_1, \ldots, z_j, \ldots, z_n)$ is such that allocates a cost for each player $j \in N$ such that:

$$\sum_{j \in N} z_j = v(N).$$

The equation above represents the "efficiency" condition in which the total cost of the grand coalition is split among its members according to z. Moreover, the "rationality" condition states that the sum of the allocated costs to each player of a coalition should be no bigger than the total cost of the coalition; e.g. no subset S of players can be formed where the cost resulting from a coalition with all of them is smaller than the total cost allocated to them according to z, namely:

$$\sum_{j \in S} z_j \leq v(S) \quad \forall S \in P.$$

In the context of a CVRPTW, a rational cost allocation vector would be one that allocates the cost to each customer such that no combination of customers can yield a smaller cost when served than the sum of their allocated values. The "individually rational condition" is a particular case for coalitions S of only one player, in which the cost allocated to it should not be bigger than its stand-alone cost: $z_j \leq v(\{j\})$. If a customer were to be assigned a cost bigger than its stand-alone cost it would not make any sense for the carriers to bring them into the pool of shared customers. Most importantly, as defined by game theory, the core in a cooperative game is the set of allocation vectors that satisfy the condition of rationality; that is:

$$\text{Core}(v) = \left\{ z \in \mathbb{R}^n : \sum_{j \in N} z_j = v(N), \sum_{j \in S} z_j \leq v(S) \quad \forall S \in P \right\}.$$

When an allocation vector is included in the core, it is said to provide "stability", given that there is no reason for players in the grand coalition to form a smaller coalition as this yields a worse outcome. The core is widely utilized as a measure of stability in research (Guajardo & Rönnqvist, 2016).

3.2.2. Shapley value allocation method

An answer to the profit(cost) allocation issue in cooperative games was introduced by Lloyd S. Shapley (Shapley, 1951, 1953). It distributes the total gains of the contributor players, assuming a non-zero contribution by all members of the coalition. The Shapley value is then calculated by accounting for all the possible orders a contributor may arrive to the coalition and computing each of their marginal contributions and average it over the total. It is based on a set of axioms which define the main characteristics a solution to the cost allocation problem should fulfill (Defryn, Sörensen, & Cornelissens, 2016). These axioms are the following:

- Symmetry: Interchangeable agents(customers) should receive the same share of the cost.
- Null player property: An agent who neither adds nor reduces the cost of any coalition should be allocated no cost.
- Efficiency: The sum of all costs allocated among all the agents should total the total coalition cost.
- Additivity: The sum of the cost allocated to agents a and b should be the same as that allocated to a third partner representing them.
- Individual rationality: The cost allocated to an agent cannot be larger than its stand-alone cost.

The allocation through the Shapley value to each player i can be calculated using 4 where the final term is the computation of the marginal contribution of player i to a certain sub-coalition S, being $v(S)$ the characteristic function which describes the worth of coalition S, e.g.: the total expected payoff S can obtain by cooperation. These marginal costs are calculated for each possible sub-coalition and then averaged by the weight that each possible combination of such sub-coalition may be reached.

$$\phi_i(N, v) = \frac{1}{N!} \sum_{S \subseteq N \setminus \{i\}} |S|!(|N| - |S| - 1)![v(S \cup \{i\}) - v(S)] \quad (4)$$

The following simple example depicts the calculation of the Shapley value. Suppose there are three players $\{1, 2, 3\}$ participating in a collaborative game with each of the possible sub-coalition resulting in the payoffs summarized in Table 3. Further, as we add players to the coalition in each of the 6 ways the grand coalition could be formed, each player in turn is going to add a marginal contribution as displayed in Table 4. Therefore, the expected marginal contribution and Shapley value for each player is simply the average value of these contributions.

In our scenario, we use the Shapley value to compute the cost each customer contributes to the total cost of serving them all together. That is, each customer is a player and we calculate the costs resulting from the objective function of the VRP for all the possible customer combinations.

Although the Shapley value defines a fair way of dividing the grand coalition's payment among its members, its result does not guarantee stability. This means some players might be better off when forming a different coalition than the grand coalition. However, the existence of a stable cost

Table 3: Payoffs of each sub-coalition for an example of a collaborative game.

S	null	{1}	{2}	{3}	{1,2}	{1,3}	{2,3}	{1,2,3}
v(S)	0	3	1	1	6	5	5	10

Table 4: Computation of the Shapley value for each player of a collaborative game.

Order	Marginal Contribution		
	1	2	3
(1,2,3)	3	3	4
(1,3,2)	3	5	2
(2,1,3)	5	1	4
(2,3,1)	5	1	4
(3,1,2)	4	5	1
(3,2,1)	5	4	1
Average	4.17	3.17	2.67

allocation is not assured and in a practical scenario, a company would only be able to accept whether to join the grand coalition or not, without any influence in the cost allocation (Vanovermeire et al., 2014). This is because the using the Shapley value method results in a solution which is "unique". Consequently, players do not have the information to challenge the allocated costs (Krajewska et al., 2008). Although, the non-emptiness of the core could be implemented as a way to verify the stability of the allocation solution and its effectiveness (Shi et al., 2020), we will not cover that aspect in our methodology. Moreover, the Shapley value derives in a cost allocation that is individually rational for a superadditive game such as the one concerning our work (Moulin, 1991).

As we have shown, the Shapley value of a participant is the expected marginal contribution of such player to a coalition picked randomly. From a practical perspective, the Shapley value has been applied in many real-world cases (Deidda et al., 2009; Littlechild & Owen, 1973; Moretti & Patrone, 2008; Moretti et al., 2007), being suggested as best practice in horizontal collaboration in logistics (Biermasz, 2012). Any cost allocation method that would steer away from the assumptions of the Shapley value method would also be valid for the application of this research, and would not modify the resulsts, to the extent that they remain a fair mechanism to divide the costs. Therefore, we leave the selection of the best allocation method outside of our work and use the Shapley value as the cost allocation mechanism for the remaining.

3.2.3. Approximation methods to the Shapley value: Random sampling

The drawback of the Shapley value being used as a cost allocation mechanism is the need of calculating the value of the characteristic function for every possible sub-coalition. This results in a very challenging computational effort that can rise exponentially to being impractical for even a low number of players. The number of possible subsets of a set consisting of n players is 2^n. Already for 30 players the amount of sub-coalitions rise over 1 billion. In particular, it is an NP-complete computing problem, therefore being too expensive computationally (Deng & Papadimitriou, 1994; Faigle & Kern, 1992).

An approximation method of the Shapley value using random sampling was first introduced by Castro et al. (2009). This procedure uses a random sample of the marginal vectors. Through sampling, statistical inference can be derived where it is impossible or impractical to obtain information for the entire population (Castro et al., 2009; Cochran, 1977). As shown by Castro et al. (2009) the estimations are efficient if the value of the characteristic function of any coalition can be calculated in polynomial time. The method consists of selecting a subset of orderings from all $n!$ possible orderings, computing the marginal contributions for all players and all orderings and approximating the Shapley value for each player by averaging the marginal contributions obtained. For the simple example introduced previously, a random sample of three orderings out of the total 6 gives the results shown in Table 5.

As expected this is very sensitive on the samples drawn for a low value of samples but becomes more accurate when increasing to a higher value of sampling (Castro et al., 2009). From the computation executed on problems of 10 customers, we conclude that with a sample size of 10 times the amount of customers an error no bigger than 3% is obtained for the computation of the Shapley value of a customer. Although this provides enough accuracy for our analysis, a better structured method is used as explained in the following section.

3.2.4. Structured Random Sampling

As shown in the example introduced and summarized in Tables 4 and 5, the calculation of the Shapley value through random sampling, yields an error which is partially due to the lack of ordering of players in the samples. Player 1 for example appears first in two of the samples and never appears last. These orderings prove to be very influential for the computing of the Shapley values.

A structured random sampling method was introduced by van Campen et al. (2018) which optimizes the random sam-

Table 5: Computation of the Shapley value through random sampling.

Order	Marginal Contribution		
	1	2	3
(1, 2, 3)	3	3	4
(1, 3, 2)	3	5	2
(2, 1, 3)	5	1	4
Average	3.67	3.00	3.33

pling method and reduces the resulting error. Their method seeks to ensure that each player is equally assigned to each position in the ordering. Therefore, each player's marginal contribution to a coalition of a defined size is calculated the same number of times and is equally distributed. To achieve this, a swapping mechanism is added to assign players in the right position in the orderings which are randomly selected. The total amount of orderings selected is divided into n subsets of length t in which each player is ordered r times in each position. The marginal contributions of the ordered player are then used to compute the Shapley value.

Considering the 3-person ($n = 3$) game example and the same $r = 3$ orderings randomly sampled before, we have therefore the size of each subset being $t = 1$. For each of the players now, each position is assigned t times in each of the subsets and their marginal contribution is calculated. This is summarized in Table 6. It can be observed that the values resulting from the use of structured random sampling are closer to the exact Shapley values than when using simple random sampling.

In their work van Campen et al. (2018) observe that, although the same number of marginal contributions per player is calculated, their method yields a small extra computational cost deriving from the swapping operation. Further, whereas random sampling is efficient, the structured method is shown by the authors not to be. Nonetheless, the structured random sampling method outperforms its simpler counterpart in terms of accuracy when it comes to approximating the Shapley values. On average the error in the Shapley value approximation is reduced by almost 30% with only a slight increase in average computation times. Consequently, we use structured random sampling in our work.

3.3. Regression analysis

For the statistical modelling part of our work, regression analysis will be conducted. Regression analysis is a set of statistical methods for determining the relationship between the dependent variable and the independent variable(s). These last are also called "explanatory" variables as the intention of the analysis is to find the degree to which each of these variables explains the behavior of the dependent one (Sykes, 1993).

Correlation is used which describes the association between two variables to the extent that a change in one is found next to a predictable change in the other. The correlation of two variables is calculated as the ratio between the covariance and the product of the standard deviations of each variable (Brown, 2016), that is:

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

The correlation can be measured through the Pearson correlation coefficient which measures the linear correlation between two sets of data. Both the strength and direction of the association between the variables are expressed through the coefficient (Yule & Kendall, 1968). A positive coefficient indicates a direct relation whereas a negative, an inverse one. The strength is measured by how close the absolute value of the coefficient is close to 1, with 0 indicating no correlation (Dowdy et al., 2011).

With multiple linear regression, several independent variables x_i are assigned to one dependent variable y, which is expressed as a linear combination of the independent ones (Rawlings et al., 1998), that is:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \epsilon$$

where parameter α represents the intercept value of y when all independent variables are 0 and β_i is the coefficient to which each variable x_i is affected to explain the most of y. The regression error ϵ is added to account for the difference between the predicted and the actual observed value of y.

When performing linear regression, the ordinary least squares (OLS) method is one of the most used. It consists of a method for defining the unknown coefficients of the independent variables in a linear regression model by minimizing the sum of the squares of the regression errors.

When dealing with multiple linear regression it is of special importance to ensure that the underlying assumptions are satisfied. The most important are the following according to Osborne and Waters (2002):

- Variables are distributed normally
- Model is linear in parameters
- No multicollinearity, e.g. low correlation between independent variables
- Homoskedasticity; e.g. the variance of ϵ remains constant for all cases of x
- No serial-correlation between error terms

Table 6: Computation of the Shapley value through structured random sampling.

Subset	Order	Swap 1	Marginal Contribution 1	Swap 2	Marginal Contribution 2	Swap 3	Marginal Contribution 3
1	(1, 2, 3)	(1, 2, 3)	3	(2, 1, 3)	1	(3, 2, 1)	1
2	(1, 3, 2)	(3, 1, 2)	4	(1, 2, 3)	3	(1, 3, 2)	2
3	(2, 1, 3)	(2, 3, 1)	5	(3, 1, 2)	5	(2, 1, 3)	4
Average			4		3		2.83

To satisfy these assumptions, generally the data and features used in the model need to be reduced. Feature selection, in which a subset of relevant predictors is selected for use in a statistical model, can be used for this purpose. All these methods are applied in the statistical analysis of our work.

4. Data generation: Instances solving and cost allocation

In this section, we dedicate ourselves to the generation of data through the application of our methodology to different problems or instances. We start by solving simple CVRP without time window constraints so as to gain insights into the basic characteristics of the problem. Further, we continue with more well-known instances with the addition of time limitations. Lastly, we define the features that we use for the approximation of the costs allocated.

4.1. Analysis of small CVRP

For the analysis of simpler collaborative single-depot capacitated vehicle problems, we focus on the modelling of instances with 3 carriers who are assigned 10 customers randomly. 10 instances are modelled with customers distributed in a squared grid of dimension 10 as the one displayed in Figure 2a. For each of the instances, all combinations of customers are modelled and solved to optimization so as to obtain the exact Shapley value assigned to each customer. This value is illustrated in Figure 2b. It is worth noting that, for these simple instances and, with a similar share of customers for each carrier, the number of customers served by each vehicle is always between 3 and 4. Therefore, there is no value at this point in the research to use this measure in our analysis. We discuss this further in later sections.

Likewise, the independent CVRP for each of the carriers is solved and their individual cost is calculated. The savings each carrier gains from the collaboration are displayed in Table 7 where the total saving is the difference between the cost of serving each of their customers and the cost when summing the Shapley value assigned for the customers brought into the coalition. These savings are all positive regardless of whether all customers in each of the carriers' operations yield positive gains. Although the grand coalition might not be the best possible coalition for each carrier, it still yields a benefit when collaborating all together. Further, as explained in Section 3.2.2, due to the individual rationality of the Shapley value, the cost of a player always results in a lower value than its stand-alone cost.

For these instances, a linear regression model is created between the Shapley value assigned to each customer and the distance to the depot as the only dependent variable. The results, summarized in Table 8, show a high degree of correlation. We avoid imposing an intercept to our function given the nature of the problem so as to start from a base cost of 0 and only depend on the distance. The regression is showcased in Figure 3 where a clear fit is observed. Moreover, heteroskedasticity can also be distinguished from the increase in the variance of the errors as the distance increase. This suggests that the measure of distance becomes less of a predictor as customers are more distant from the depot where, given a bigger spread between customers, the limited capacity of the vehicles could play a bigger role. We deal with the problem of heteroskedasticity with robust regressions in the remaining of the work.

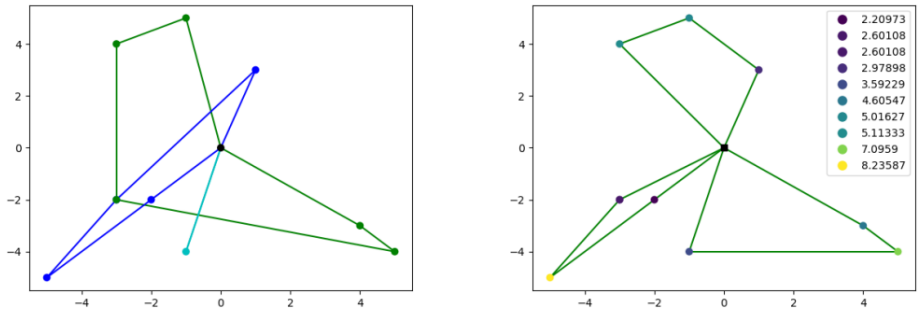
Finally, the result of the cost allocation is summarized in Figure 4 where the average cost for customers located in different placements is represented. It is noticeable how a simple rule based on distance can be applied to the solving of a much more complex logistics problem.

4.2. Solomon instances

In order to be able to benchmark the results obtained for the CVRPTW and ensure a correct implementation of the structured random sampling algorithm used for the allocation of costs, the use of well-known data instances is implemented, the Solomon instances (Solomon, 1987). Moreover, this allows for the further study of our results in consequent works.

Solomon instances are divided into different sets according to their characteristics. R and C sets in the data differ in that the R sets (R1 and R2) have the geographical data randomly generated, whereas the C sets (C1 and C2) are set in geographical clusters. For the RC1 and RC2 sets, a mix of random and clustered structures is used. The numbering relates to the limitation on the scheduling horizon, having smaller time windows and allowing a few customers per route in sets 1 and a long scheduling horizon with more customers in sets 2. The customer locations remain the same within all problems of one type (i.e., R, C and RC), mainly differing in the tightness of their time windows, that is the time window width. There exists also variance within the time window density, that is the percentage of customers with tight time windows, ranging from 25% to 100% (Solomon, 1987).

Solomon introduced problems of 100 customers where travel times match the euclidean distances between the ge-



(a) CVRP without collaboration of 3 partners (green, cyan and blue).

(b) CVRP with collaboration of 3 partners. Shapley value for each customer in color.

Figure 2: Instance of a collaborative vehicle routing problem with 3 partners and 10 customers.

Table 7: Total savings obtained for each carrier (A, B, C) when collaborating in a CVRP.

Instance	A	B	C
0	4.77	4.89	2.84
1	1.22	5.48	8.53
2	1.82	4.88	0.88
3	3.65	7.29	1.08
4	3.47	8.21	7.71
5	2.47	6.73	8.43
6	4.07	5.77	8.56
7	7.49	4.38	5.60
8	9.42	7.94	5.19
9	2.78	5.50	3.00
10	1.72	7.73	4.42

Table 8: Linear regression model summary with the cost per customer as the dependent variable and distance as the sole independent variable.

	coef	std err	t	$P > \|t\|$	[0.025	0.975]
distance	1.0162	0.02	51.11	0	0.977	1.056

ographical points. For computational reasons, we base our work on the instances of the 50 first customers in each of them where 100% of customers are assigned a time window constraint. The instances used in our analysis and their characteristics are summarized in Table 9.

Most importantly, the design of the instances highlights aims at addressing several variables that generally influence the routing and scheduling when solving CVRPTW problems such as geographical location, vehicle capacity, amount of time-constrained customers, and tightness and positioning of the time windows assigned. (Solomon, 1987).

4.3. Cost function: features and fit

With the aim of developing our cost function, different statistical models are applied to the data generated out of allocating the cost via our method using the Shapley value for each customer in Solomon's 50 customers instances. An example of such data can be observed in Figures 5 and 6 where the solution for instances R101.50 and R201.50 respectively are displayed. The vehicle routes are visible in conjunction with the computed Shapley values for each of the customers.

In the first example, having time windows which are more constraining, the vehicles route less customers in smaller loops. In contrast, in the latter, routes are more convoluted due to time constraints being more relaxed. This is also reflected in the spread of Shapley value which has a bigger variance and seems to depend more on distance in the first case compared to the second one. This example signals that short scheduling horizons, allowing for only a few customers to be serviced per route, yields a higher variance of the cost allocated for each customer.

An approximation to the cost function should try to resemble the real interactions of the CVRPTW as much as possible. For this reason, the definition of certain features should match the factors that affect this problem as discussed previously in Section 4.2. Therefore, different features have been

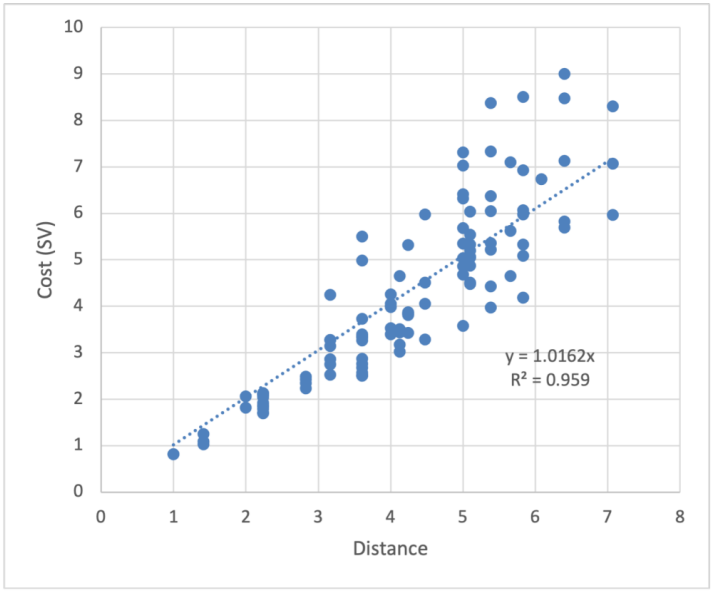


Figure 3: Relation between the cost allocated and the distance to the depot for simple CVRPs.

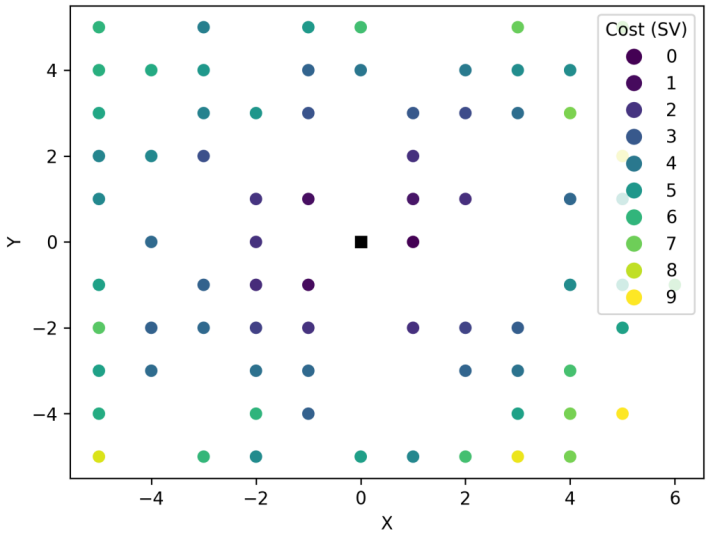


Figure 4: Cost allocated for customers in simple CVRPs.

defined following the three main dimensions that affect the cost allocation problem. In order to generalize their application and derive clearer conclusions from their coefficients, relative measures are used when possible.

Capacity-based features
The main parameter affecting the usage of vehicles is the demand of customers and the capacity of the vehicles. In order to represent the first effect in our function we introduce a feature relating the demand of each customer to the average demand of the population, that is:

$$\mathrm{CF}_1^i = \frac{d_i}{\frac{\sum_{i \in N} d_i}{|N|}}$$

Further the capacity of vehicles is reflected in the demand related to the vehicle capacity, e.g.:

$$\mathrm{CF}_2^i = \frac{d_i}{C}$$

where C is the vehicles' capacity.

Distance-based features
As observed in our modeled examples before, geographical distance appears as a main variable to take into consideration. Firstly, we introduce a feature comparing the distance of a customer i to the depot $i = 0$:

$$\mathrm{DF}_1^i = \frac{c_{i0}}{\frac{\sum_{i \in N} c_{i0}}{|N|}}$$

Table 9: Solomon instances used for the analysis and their characteristics.

Instance	Amount of Vehicles	Vehicles' Capacity	Average Demand	Average distance to depot	Average length of Time Windows	Service Time	Amount of customers w/ Time Window
C101.50	25	200	17.20	24.11	60.14	90	50
C201.50	25	700	17.20	26.79	160.00	90	50
R101.50	25	200	14.42	26.25	10.00	10	50
R201.50	25	1000	14.42	26.25	116.46	10	50
RC101.50	25	200	19.40	40.60	30.00	10	50
RC201.50	25	1000	19.40	40.60	120.00	10	50
C108.50	25	200	17.20	24.11	240.78	90	50
C208.50	25	700	17.20	26.79	640.00	90	50
R109.50	25	200	14.42	26.25	58.94	10	50
R209.50	25	1000	14.42	26.25	351.08	10	50
RC108.50	25	200	19.40	40.60	111.62	10	50
RC208.50	25	1000	19.40	40.60	472.90	10	50

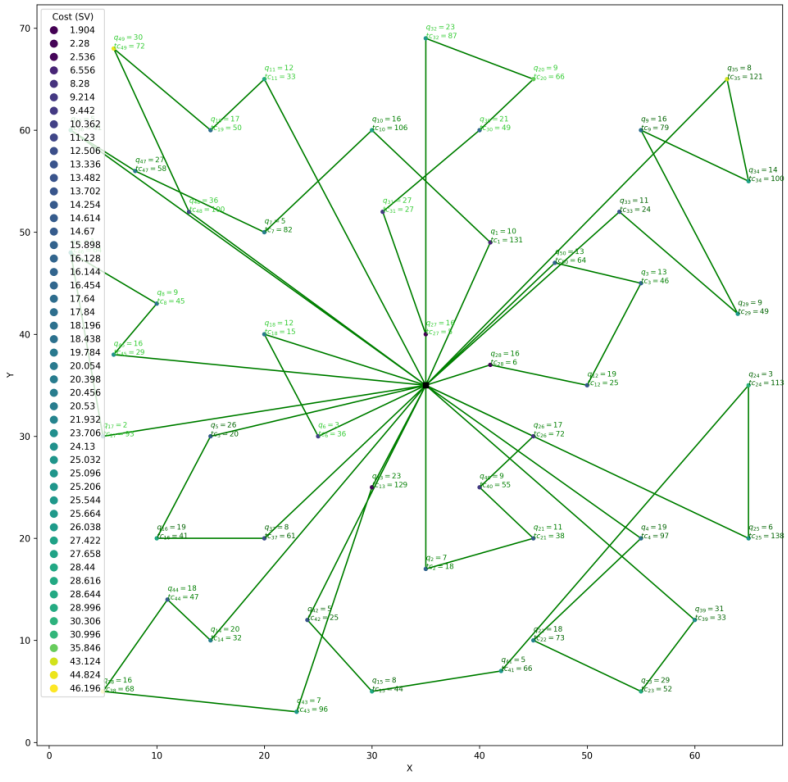


Figure 5: Solution to Solomon's instance R101.50 showing the computed cost (Shapley values) for each customer.

To note is that as all costs are taken using euclidean distances, the cost matrix is symmetrical, e.g.: $c_{ij} = c_{ji}$. Secondly, a measure of the degree of clustering is needed. Therefore the proximity to other customers is accounted for by measuring the number of points that fall within a certain distance. For our measure we use the average distance to the depot as the threshold for counting the near customers as shown in Figure 7. That is:

$$\mathrm{DF}_2^i = \|\{j \in J\}\| : d_{i,j} \in \left[0, \frac{\sum_{i \in N} c_{i0}}{|N|}\right)$$

Time-based features

In the case of the time dimension, the measure of the impact of the time windows' width is needed. With this objective, we introduce a feature in which we calculate the length of the time window as the difference between the earlier and latest service times and divide that by the average distance of customers to the depot.

$$\mathrm{TF}_1^i = \frac{b_i - a_i}{\frac{\sum_{i \in N} c_{i0}}{|N|}}$$

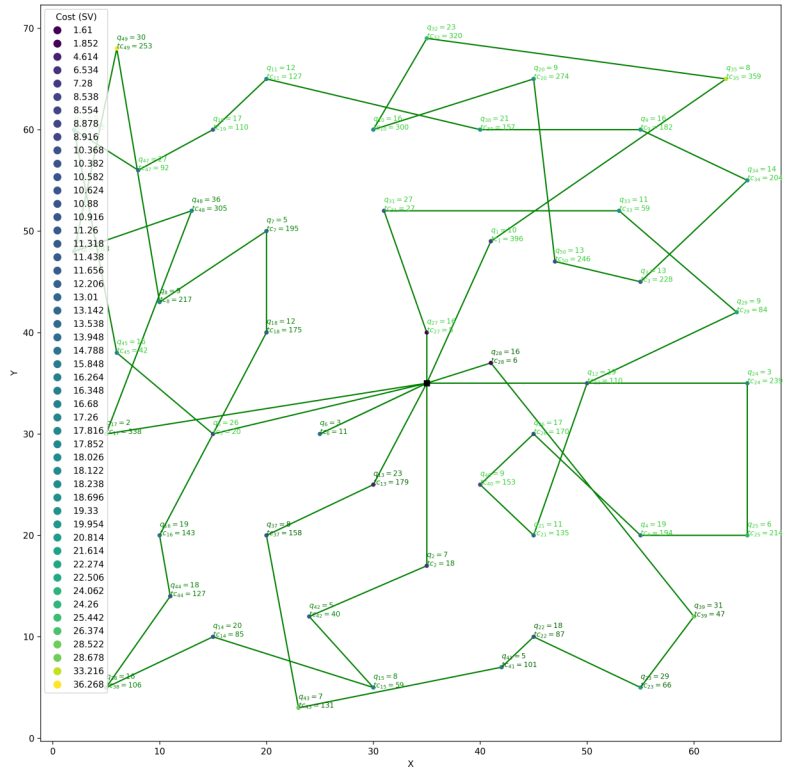


Figure 6: Solution to Solomon's instance R201.50 showing the computed cost (Shapley value) for each customer.

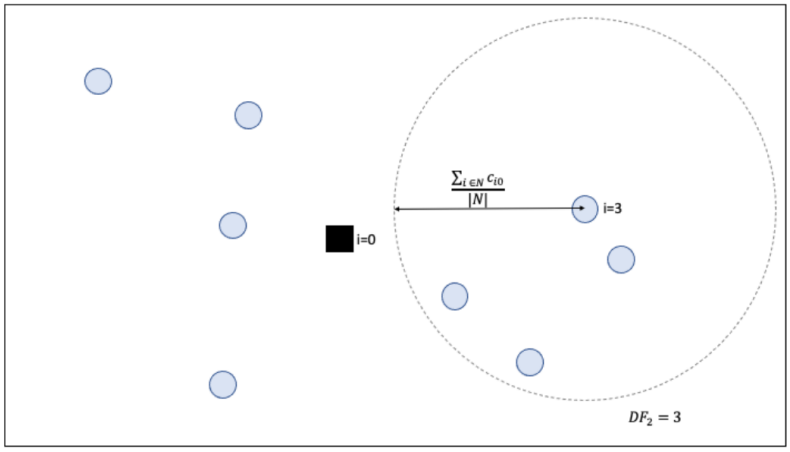


Figure 7: Example of computing the proximity customers for a customer i.

Further, as in the case of the distance between customers, the spread in the time windows is looked to be reflected in the time window overlap as defined in the time feature:

$$\mathrm{TF}_2^i = \|\{j \in J\}\| : (a_i, b_i) \cap \left(a_j, b_j\right) \neq \emptyset$$

Instance-specific features

Relying on the characteristics of Solomon instances which aim at characterizing the different dimensions affecting CVRPTW, we use certain instance-specific features which aid in the analysis of the costs and their distributions between instances themselves. For this purpose we separate instances under the following classification:

- Customers proximity: Clustered/Non-clustered customers
- Time: Tight/Non-tight Time Windows

Additionally, different modelled instances are compared using other characteristics which are not binary to the extent that they are not directly related to the Solomon instances' design criteria such as:

- Average customer demand

- Ratio of total demand served by vehicle capacity
- Average length of time windows
- Average number of customers per tour

5. Analysis and discussion

With the data generated for the instances used and the features calculated we follow with the statistical analysis in the following sections.

5.1. Results obtained and observations

The results obtained from applying our methodology are compiled in Appendix 1 where the characteristics of each node from Solomon instances in conjunction with the cost assigned in the grand coalition are shown.

As in Section 4.1, we begin by addressing the total savings of each carrier when collaborating which are summarized in Table 10 where the average value has been added for further study. We can again observe that all gains are positive, in line with the research findings explained in Section 2.1, and further enlarging the case for collaboration between carriers. Worth noting is that this is the total gains for each carrier (each of which has a similar share of assigned customers), which does not exclude the possibility of a customer bringing a negative gain when shared into the common pool of the joint operation. This is later explored in Section 5.4.

We can therefore observe the following:

Observation 1. The total savings calculated for each individual carrier when collaborating in a grand coalition are found to be positive in all instances.

Within this level of granularity, we want to evaluate the differences in the costs allocated and the corresponding savings obtained for the customers in each instance. Figure 8 depicts the average values obtained in instances split by clustering and the time window horizon. In Figure 8a we observe that the clustering of customers yields a smaller average cost per customer. Further, Figure 8b shows how having short time windows affect considerably the average cost of serving the customers. Interestingly, in both cases the savings do not show the same level of change.

To further study the instance-related features, Figure 9 compiles the same measures for average cost and savings compared to the different characteristics of the Solomon instances. Firstly, an increase in the average customer demand leads to an increase in the savings as shown in Figure 9a. Moreover, when analysing Figure 9b we observe no general trend for lower demands per truck capacity but notice an increase in the savings and costs when increasing past a certain threshold. We can relate the results from Figure 8b with Figure 9c where we observe that a decrease in the average time window length results in a smaller average cost allocated to each customer. Finally, with regard to the average number of customers per tour, Figure 9d shows that average costs decrease when this increases.

Throughout the results showed, a clear relation between costs and savings can be observed, pointing towards the fact that the same limitations that increase the cost of serving the pool of customers are the ones that can be avoided when collaborating.

Let us summarize the main observations in the following:

Observation 2. Customer clustering and relaxation of time windows yield a decrease in the average cost of serving each customer.

Observation 3. An increase in customer demand or a decrease in the vehicles' capacity leads to an increase in the costs associated with each customer.

Observation 4. As the total amount of tours to serve a certain amount of customers decreases, the average cost associated with each customer also does.

Observation 5. There is a strong correlation between the costs assigned to each customer when serving them in a stand-alone operation and the potential savings when acting in a collaborative environment.

Although always dealing with a collaborative scenario, the cost allocation deriving from the joint operation can be analyzed from a non-collaborative perspective. The grand coalition acts as a shared resources all-information central planning body. For the approximation of the cost allocation with the features defined, we start with the results obtained from the centrally planned cost allocation perspective. Later, we analyze the data resulting from expanding on the collaborative nature of the problem.

5.2. Model fit and feature selection

Our main aim is to conduct a model selection procedure to choose the features and interactions of them which yield the best fit to the cost allocated to customers through linear regression. The measure used to compare the performance of different models is under debate, which questions the "best" fit (Hansen & Yu, 2001). Nevertheless, we will focus on the Akaike (AIC) and Bayesian Information Criterion (BIC) and the adjusted R^2 value as means of evaluating our models. Furthermore, for simplicity of our solution and following the principle of parsimony, or Occam's razor, we will focus on finding the model with the least amount of features and interactions given similar explanatory power.

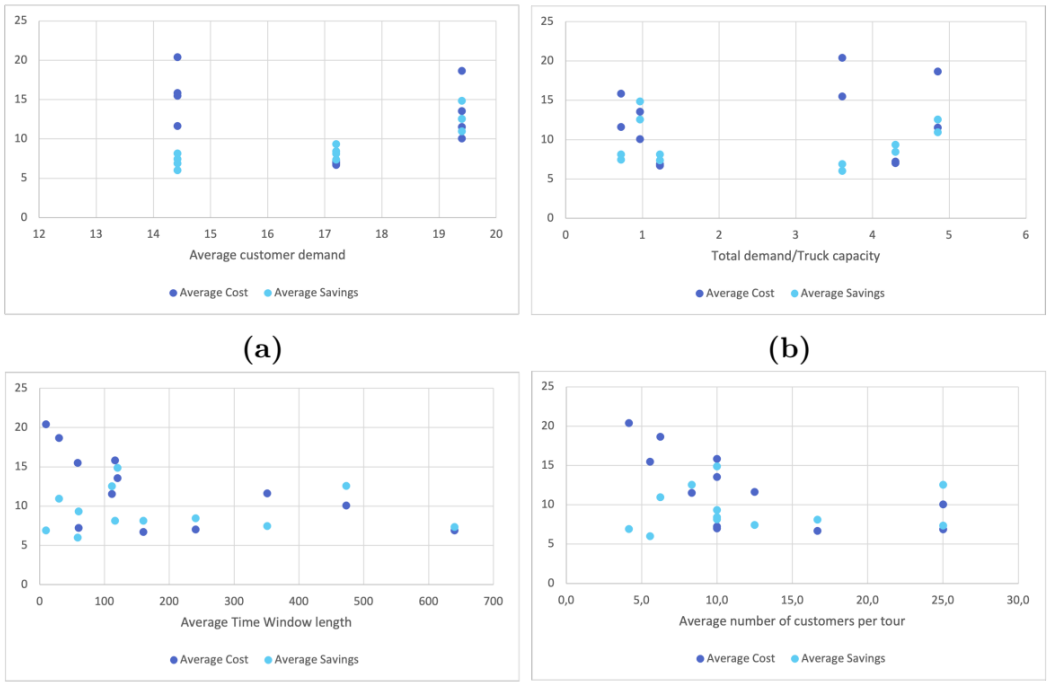
To start, a multiple ordinary least squares linear regression analysis is conducted using all features in the model as the independent variables and the calculated allocated cost (Shapley value) as the dependent one. Said analysis is performed with the data for all the combined instances shown in Table 9. At first, we conduct our analysis without including an intercept in the model, with the results being presented in Figure 10. By applying a Breusch-Pagan test to the data, the null hypothesis of homoskedasticity is discarded therefore the regression results show the scores for t-tests using heteroskedasticity-robust standard errors. The model shows a high correlation with a value of R^2 of 0.855. Regarding the distance metrics, from the values of the regression coefficients, we can infer both a positive statistically significant

Table 10: Total savings for each carrier (A, B, C) in the modelled CVRPTW Solomon instances.

Instance	A	B	C	Average
C101.50	150.75	161.57	154.59	155.64
C201.50	135.59	113.51	156.80	135.30
R101.50	120.20	123.43	101.63	115.09
R201.50	145.82	131.90	128.99	135.57
RC101.50	228.59	152.04	166.80	182.48
RC201.50	263.92	251.12	228.17	247.74
C108.50	133.97	151.41	136.62	140.66
C208.50	127.60	102.38	137.47	122.49
R109.50	78.19	92.84	129.13	100.05
R209.50	112.12	124.45	136.17	124.25
RC108.50	219.75	204.91	202.62	209.09
RC208.50	210.48	213.47	204.14	209.36

(a) (b)

Figure 8: Average costs and gains for the modelled Solomon instances split by clustering and TW horizon.



(a) (b)

(c) (d)

Figure 9: Average costs and savings for the modelled Solomon instances in relation to different instance characteristics.

correlation between the cost and the relative distance to the depot (DF1) and a negative one with the amount of proximity points (DF2). This confirms the results previously discussed in Section 5.1 in Observation 2. With regard to the time features, the negative correlation of the relative time window length (TF1) signals a decrease in the cost when having less constraining time windows which also is in line with the previous observations. Further, a positive coefficient for the time window overlap (TF2) would be against the common sense of pricing time windows and will be explained through further analysis of different models. None of the capacity constraints appear to have any statistical significance given by the p-values of capacity-related features which we analyze later.

To further explore the correlation between the variables let us analyze the relationship between each of them. Figure 11 shows a scatter plot for each of the variable pairs in the regression model. A grid of Axes is displayed such that each variable in the data will be shared across the y-axes across a single row and the x-axes across a single column. The diagonal plots show a distribution plot to represent the marginal distribution of each variable. Through an inspection of the scatter plots for the features in each dimension a clear correlation can be observed, in particular within distance-dependent and time-dependent features. This is further confirmed when calculating the correlation values as shown in Figure 12. As observed, the correlation between DF1 and DF2 is negative which means that having more points in the vicinity relates to a decrease in the distance to the depot or vice versa. Further, the correlation between features TF1 and TF2 is positive corresponding to a increase in the time window overlap when increasing the gap between the beginning and end of the allowed time for service at the customer.

As observed in the relation between the features in each dimension, and due to over-redundancy, the results obtained signal the need of relying on one main factor for each of the dimensions specified (Capacity, Distance, Time). A model selection is then conducted to choose the best fit with the measures specified before. All the combinations of features are tested and their AIC, BIC and adjusted R^2 values are calculated. It is found that features DF1, TF1 and CF2 are the ones that yield a better explanation of the dependent variable remaining one of the relatively lowest models regarding the AIC and BIC scores. It is reasonable to believe that the inclusion of CF2 is related to the addition of the maximal vehicle capacity as a variable to the model. Further, of the proximity measures included for distance and time windows none remain as the selected by the model. As we see later, this is very much related to the degree customers are clustered and time-constrained. The results for the regression with only these features are showcased in Table 11. The feature CF2 appears to have less statistical significance but has a negative coefficient which is in line with Observation 2. It is worth noting that the addition of an intercept in our model yields a lower explanatory effect over the independent variable given its R^2 value but has a higher statistical significance for all the variables including the intercept and achieves a lower AIC and BIC value as displayed in the regression results in Figure 13. The addition of an intercept would signify a base cost allocated to customers in any case which would be a simple base pricing strategy from which to start.

From the two linear regression models, with or without intercept, we can derive two main strategies to price customers. Firstly, one where no base price is defined and through the implementation of features pricing the dimensions which define each customer (demand, time window, distance) the final cost is greatly described. Secondly, assuming a basic price for all customers (intercept) and pricing the features differently resulting in reduced descriptive power of each feature on the final cost allocated to each customer. We will further study the pricing of the different features in the following section.

As a whole, from all our regression models we can observe the following:

Observation 6. The cost allocated to customers in vehicle routing problems with time windows can be estimated with a few defined features related to customers' location, time window, demand and vehicle capacity.

5.3. Customer pricing

The practical application of our work is focused on being able to price each of the characteristics that comprise the logistics problem of serving different customers whose geographical location, time demands and capacity vary. As defined in our cost-approximation function, the features introduced allow us to represent each of these dimensions in our model. None-withstanding, we want to study different ways of approaching these characteristics. As addressed in Section 4.2 the variety in the Solomon instances allow for the assessment of the results of our cost approximation when dealing both with clustered customers vs non-clustered ones and the time horizon restrictions. In this regard, we will study the differences in our regression models when dealing with each condition.

5.3.1. Clustered vs non-clustered customers

For this comparison, instances with non-clustered customers which are located geographically in a random fashion (R101.50, R201.50, R109.50, R209.50) are compared with their clustered counterparts in which customers are in groups of close proximity to each other (C101.50, C201.50, C108.50, C208.50). Instances RC, being a mix of both, are left outside of the analysis. The results of the linear regression for both cases are compiled in Tables 12 and 13 where CF1 has been dropped for being highly correlated to CF2. The explanatory power of the models gives an adjusted R^2 value of 0.624 for the clustered customers and 0.702 for the randomly located ones. It can be observed that for the non-clustered customers the intercept for the cost is not statistically relevant with only the relative distance to the warehouse being the relevant distance feature. In contrast, when dealing with instances of clustered customers, the intercept

OLS Regression Results

Dep. Variable:	Cost	R-squared (uncentered):	0.855
Model:	OLS	Adj. R-squared (uncentered):	0.853
Method:	Least Squares	F-statistic:	686.2
		Prob (F-statistic):	2.63e-263
		Log-Likelihood:	-1859.1
No. Observations:	600	AIC:	3730.
Df Residuals:	594	BIC:	3757.
Df Model:	6		
Covariance Type:	HC1		

	coef	std err	z	P>|z|	[0.025	0.975]
DF1	14.7125	0.673	21.848	0.000	13.393	16.032
DF2	-0.0542	0.023	-2.394	0.017	-0.099	-0.010
CF1	0.4697	0.503	0.933	0.351	-0.517	1.456
CF2	-10.7873	6.737	-1.601	0.109	-23.993	2.418
TF1	-0.3918	0.035	-11.219	0.000	-0.460	-0.323
TF2	0.0334	0.014	2.424	0.015	0.006	0.060

Omnibus:	45.913	Durbin-Watson:	0.952
Prob(Omnibus):	0.000	Jarque-Bera (JB):	64.946
Skew:	0.589	Prob(JB):	7.89e-15
Kurtosis:	4.100	Cond. No.	936.

Figure 10: Ordinate least squares regression results for all instances with all features in the model.

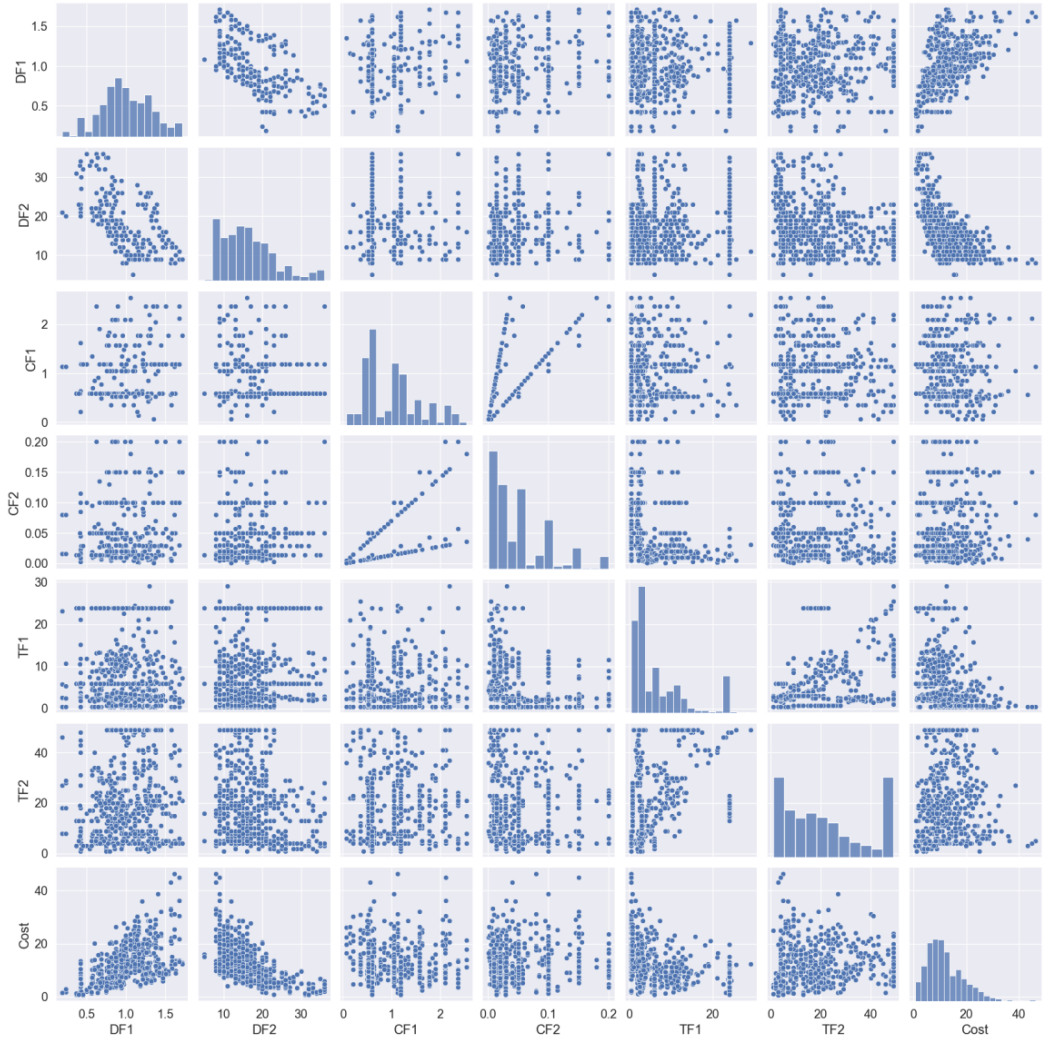


Figure 11: Scatter plots of all the pairs of variables in the model. The variables' distributions are shown in the diagonal plots.

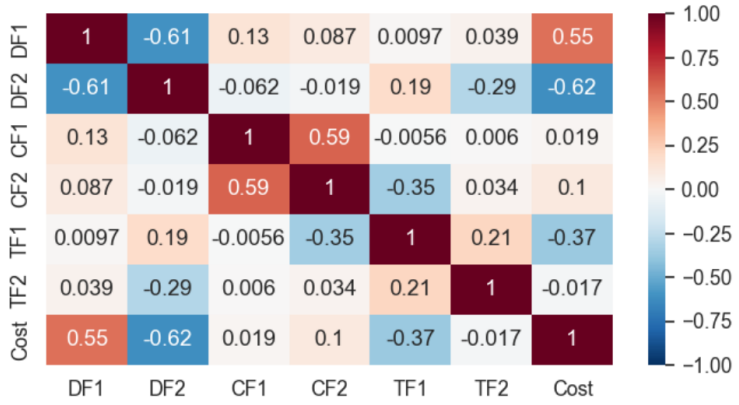


Figure 12: Correlation values between each feature in the model.

Table 11: Linear regression model summary with the cost (Shapley value) as the dependent variable and the selected features as independent variables.

	coef	std err	z	$P > \|z\|$	[0.025	0.975]
DF1	14.933	0.492	30.380	0.000	13.970	15.897
TF1	-0.388	0.033	-11.893	0.000	-0.452	-0.324
CF2	-8.322	4.829	-1.723	0.085	-17.787	1.143

OLS Regression Results

Dep. Variable:	Cost	R-squared:	0.449
Model:	OLS	Adj. R-squared:	0.447
Method:	Least Squares	F-statistic:	106.8
		Prob (F-statistic):	2.43e-55
		Log-Likelihood:	-1859.3
No. Observations:	600	AIC:	3727.
Df Residuals:	596	BIC:	3744.
Df Model:	3		
Covariance Type:	HC1		

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	2.6905	0.790	3.405	0.001	1.142	4.239
DF1	12.9943	0.854	15.221	0.000	11.321	14.667
TF1	-0.4278	0.031	-14.011	0.000	-0.488	-0.368
CF2	-13.7909	5.104	-2.702	0.007	-23.794	-3.788

Omnibus:	58.953	Durbin-Watson:	0.977
Prob(Omnibus):	0.000	Jarque-Bera (JB):	90.781
Skew:	0.682	Prob(JB):	1.94e-20
Kurtosis:	4.331	Cond. No.	219.

Figure 13: Regression results for the selected model features with an added intercept.

becomes relevant in conjunction with the two distance-based features. In particular, the addition of DF2 which considers the proximity to other customers affects the cost negatively. Further, for the randomly located customers, the time features related to the length and overlap of time windows appear to be statistically significant and affect both the cost negatively. This would signify that time windows have a major impact on the cost of serving customers when these are not grouped in clusters of close proximity. Moreover, the difference in the coefficient for DF1 of more than 10 signifies a significantly bigger impact on the distances in the case of non-clustered customers compared to the clustered ones. All in all, our results clearly show how the grouping of customers in clusters can have a clear effect on the variables which mainly affect the allocation of costs.

Further, having the geographical location as the only feature we would like to assess the pricing of certain regions or locations for our customers. In this regard, we compile all the customers from the instances mentioned above so as to show how distance to the depot, which translates into time in our model, affects cost allocation. Figure 15 shows the case of customers with random locations where zoning related to the distance to the warehouse is applied. In the case of clustered customers we summarize the customers of in-

Table 12: Linear regression results for the non-clustered customers.

	coef	std err	z	$P > \|z\|$	[0.025	0.975]
const	10.011	5.494	1.822	0.068	-0.757	20.779
DF1	14.666	2.351	6.238	0.000	10.058	19.274
DF2	-0.282	0.206	-1.366	0.172	-0.687	0.123
CF2	-4.769	9.865	-0.483	0.629	-24.104	14.566
TF1	-0.328	0.076	-4.296	0.000	-0.477	-0.178
TF2	-0.121	0.030	-4.057	0.000	-0.180	-0.063

Table 13: Linear regression results for the clustered customers.

	coef	std err	z	$P > \|z\|$	[0.025	0.975]
const	7.230	1.360	5.318	0.000	4.565	9.895
DF1	4.272	0.537	7.952	0.000	3.219	5.325
DF2	-0.215	0.036	-6.019	0.000	-0.285	-0.145
CF2	1.374	3.472	0.396	0.692	-5.431	8.179
TF1	-0.033	0.030	-1.121	0.262	-0.091	0.025
TF2	0.029	0.024	1.201	0.230	-0.018	0.077

stances RC101.50, RC201.50, RC108.50 and RC 208.50 in their respective clusters and then calculate an average cost per cluster which is represented in Figure 14. Both figures exemplify how different zoning rules can be applied to better adapt to the different structures of customer locations.

5.3.2. Tightness of time windows

As in the case of distance, we proceed to analyze the effect of time in our problem by comparing instances which have a shorter time-horizon (e.g. R/C/RC 101.50) to those which a longer one (e.g. R/C/RC/201.50). For the case with tighter time restrictions, it can be observed that the significant features are only the relative distance to the depot (DF1) and both time features. The limitation that the time constraint produces, makes the overlap of time windows (TF2) a significant feature in contrast with the case with longer time horizons. Moreover, in the case with more relaxed time constraints the intercept becomes statistically significant yielding a basic cost as a base from which then the first distance measure and the relative length of the time windows both add and subtract respectively to the cost allocated for each customer. Further, the clustering of customers has a statistically significant correlation in this scenario as is the case with the capacity.

In general, we observe that short scheduling horizons, allowing for only a few customers to be serviced per route, yield a higher variance of the cost allocated to each customer.

To summarize, the individual regression models of each of the different set of instances has shown how the pricing strategy of customers depends greatly on the configuration of them. It can be derived that the classification of customer configurations among the categories used results in a better understanding of how to divide the costs when serving them.

5.4. Collaborative gains analysis

One of the main areas of focus of our work is the cooperation aspect of the logistics carriers in this problem. Given that we are dealing with a collaborative scenario, we are interested in evaluating the allocation of costs of each carrier when operating individually compared to that when in cooperation. So far, we have allocated costs based on a joint logistic operation in which we have assumed that all carriers share their entire pool of customers and vehicle resources so as to achieve the most efficient service from a logistics perspective. There is however a reluctance of carriers to share all their customers' information given that this is from where their business extracts great value.

As explained, for each of the instances modelled we have assigned each customer in the pool to one of three carriers. We therefore utilize our methodology within each carrier's operation solving the CVRPTW and using structured random sampling to calculate the Shapley value and allocate the cost of each customer in a partner stand-alone operation. These results can be found in Appendix 1 under the carrier's assignment and the stand-alone cost.

Moreover, for each of the customers of the competitors, a cost is calculated considering what would be the cost allocated to that customer if the entire operation was to be delivered to an identical pool of customers with the only addition of this extra one. This is represented in Figure 16. In this scenario, the mechanism of an auction is being reproduced where a carrier would offer a customer to the pool of shared customers. In 5.9a carrier 1 would offer a customer for which it has a high cost allocated. Secondly, in 5.9b the other carriers calculate their cost when adding this customer to their service. Finally, the carrier with the lowest servicing cost wins the auction given that he can pay a higher price because he can include the customer with the lower marginal cost. This results in the transfer of this customer from one operator to the other. Considering this, we calculate for each

Cluster	Distance to Depot	Cost
1	37.8	12.2
2	36.5	11.5
3	36.1	12.4
4	40.8	13.0
5	51.0	18.1

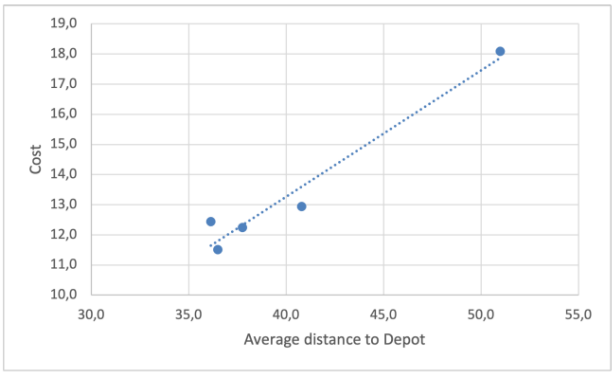


Figure 14: Cost assigned for each customer cluster in clustered instances.

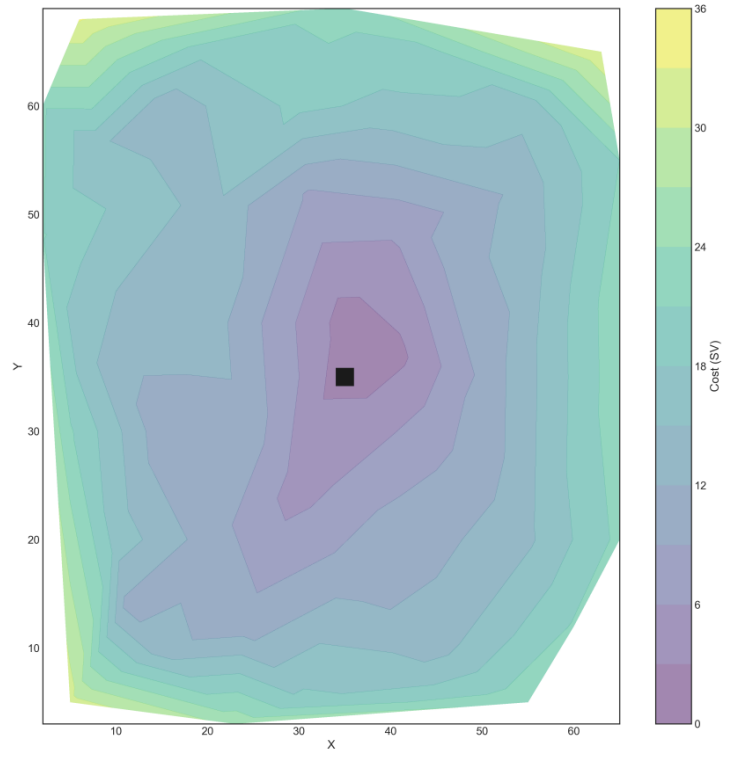


Figure 15: Distribution of cost for customers in instances with randomly chosen locations.

Table 14: Linear regression results for the instances with a shorter time horizon.

	coef	std err	z	$P > \|z\|$	[0.025	0.975]
const	5.490	3.071	1.788	0.074	-0.529	11.508
DF1	16.134	2.187	7.378	0.000	11.847	20.420
DF2	0.029	0.087	0.335	0.737	-0.141	0.200
CF2	-16.001	8.931	-1.792	0.073	-33.505	1.504
TF1	-5.495	0.566	-9.712	0.000	-6.604	-4.386
TF2	0.126	0.057	2.191	0.028	0.013	0.239

carrier the Shapley value for each customer of the competitors. The lowest cost from the computation of other carriers' operations is presented in the last column in the table found in Appendix 1. Our calculations are not completely representative of the entire auctioning process because they only consider the first iteration of it. This proves to be sufficient for our analysis.

We follow with the analysis of two scenarios. In the first

Table 15: Linear regression results for the instances with a longer time horizon.

	coef	std err	z	$P > \|z\|$	[0.025	0.975]
const	17.964	2.938	6.115	0.000	12.207	23.722
DF1	7.918	1.593	4.969	0.000	4.795	11.041
DF2	-0.402	0.067	-5.972	0.000	-0.533	-0.270
CF2	-98.750	35.031	-2.819	0.005	-167.410	-30.091
TF1	-1.096	0.418	-2.622	0.009	-1.915	-0.277
TF2	-0.042	0.090	-0.470	0.638	-0.219	0.134

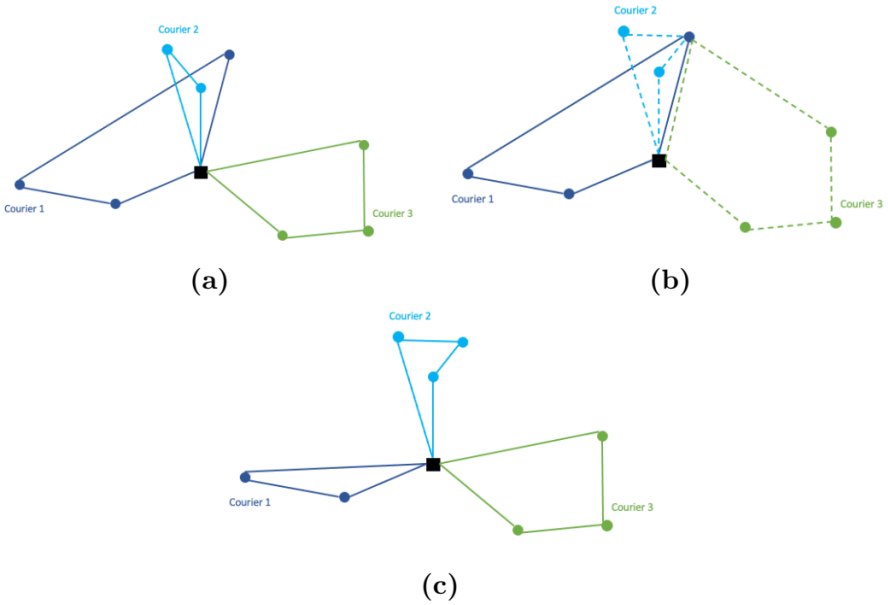


Figure 16: Representation of an auction mechanism for the sharing of customers in a collaborative vehicle routing problem.

one, we compare the stand-alone operation of each carrier with the grand coalition (central plan) where all customers are shared. For the second scenario, we consider the aforementioned auction mechanism in which we compare the cost of every customer in each carriers' operation to that of their competitors.

5.4.1. Central plan comparison

The aim is to derive insights into which customers become detrimental when being shared which defeats the purpose of sharing the information in the search of reducing costs. Accordingly, we calculate customers' Shapley value in carriers' stand-alone operation and in a grand coalition. We then classify customers who yield a negative/positive gaining so as to perform a logistic regression with the variables in the data. A dummy variable is used with a value of 1 for customers with a positive gain when shared and 0 otherwise. The aim is to have a prediction of which customers should be shared and which should be kept. Consistently with the rest of our work, the instances mentioned in Section 4.2 are used. Of the 600 customers, 33 yield a negative gain. From our statistical analysis, only the relative distance yields any statistical significance in the explanation of the dependent variable. The results for the regression when using only this feature are condensed in Figure 17.

Further, the gains for each customer when cooperating are also regressed with the features constructed so far. The results are showcased in Table 16 where the effect of both distance features are found to have a significant correlation.

Apart from the significant descriptive value of the distance features, there is no particular difference found for the 33 customers for which gains are negative. This constitutes a paradox, given that the cost for these customers is higher in the collaborative scenario than when in a partner stand-alone undertaking, but still carriers benefit from the collaboration as a whole when summing all customers. There is, therefore, a contradiction between the cost-allocating problem and the logistics problem which represents an area of potential future work.

In general, we found a very strong correlation between the costs allocated to customers when carriers act alone and those when doing so in a collaborative scenario. Figure 18 shows this relationship which yields a high value of R^2, confirming Observation 5.

5.4.2. Auction scenario

Analogously, the same procedure from the central plan scenario is applied to the auction one. Of the 600 customers

Logit Regression Results

Dep. Variable:	sell_customer	No. Observations:	600
Model:	Logit	Df Residuals:	599
Method:	MLE	Df Model:	0
		Pseudo R-squ.:	-0.01568
		Log-Likelihood:	-129.79
converged:	True	LL-Null:	-127.79
Covariance Type:	nonrobust	LLR p-value:	nan

	coef	std err	z	P>|z|	[0.025	0.975]
DF1	3.0327	0.207	14.621	0.000	2.626	3.439

.....

Figure 17: Logistic regression result when predicting a customer's gain in collaboration.

Table 16: Linear regression results for the gains obtained for each customers when carriers collaborate.

	coef	std err	z	$P > \|z\|$	[0.025	0.975]
const	8.258	1.988	4.153	0.000	4.361	12.155
DF1	6.570	1.153	5.700	0.000	4.311	8.830
DF2	-0.370	0.062	-5.931	0.000	-0.492	-0.247
CF2	11.034	9.540	1.157	0.247	-7.664	29.732
TF1	0.044	0.043	1.017	0.309	-0.040	0.127
TF2	-0.007	0.021	-0.328	0.743	-0.047	0.034

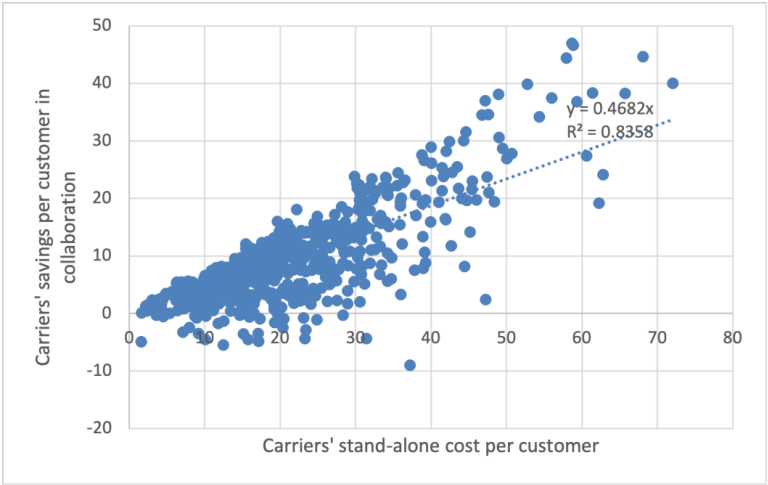


Figure 18: Correlation between savings in collaboration and costs in stand-alone allocated to customers.

studied, 167 yield a negative gain when comparing it to the lowest servicing cost from other carriers; that is, only 167 customers are serviced with the lowest cost by the carrier they have been assigned to. We follow with the same logistic regression as with the case before which only has a significant correlation with DF1 with a coefficient of 0.90. Moreover, a linear regression model is constructed with the defined features as independent variables and the gains by auctioning as the dependent one. The results from the OLS regression is shown in Table 17. Although having a low explanatory power with an adjusted R^2 value of 0.069 , the intercept has a significant statistical value. Noticeably, the only other feature with statistical significance is DF2 measuring the degree of clustering of customers. This represents that customers which are closer to other customers yield a lower cost for other carriers than the one serving them.

Given the outcome observed for this scenario, it is sensible to expect all total savings of carriers when collaborating to be positive, as our results in Section 5.1. This is explained in the fact that the great majority of customers can be serviced with lesser cost by a different carrier than the one to which they have been assigned. In practical scenarios, decentral collaboration can be exercised for the sharing of selective customers which are found to bring the biggest cost. In our results, we find that the average cost allocated in a standalone operation to customers that are cheaper to service by other carriers is 23.84 compared to 15.32 for those which are not. This represents a significant difference of 56% bigger cost.

Table 17: Linear regression results for cost of customers compared to other carriers' cost.

	coef	std err	z	$P > \|z\|$	[0.025	0.975]
const	7.561	2.288	3.305	0.001	3.077	12.045
DF1	1.669	1.371	1.218	0.223	-1.018	4.356
DF2	-0.292	0.069	-4.213	0.000	-0.428	-0.156
CF2	15.911	9.921	1.604	0.109	-3.533	35.355
TF1	0.011	0.047	0.236	0.813	-0.082	0.104
TF2	-0.027	0.023	-1.179	0.238	-0.071	0.018

6. Conclusion and recommendations

This study explores how cost allocation to customers served in Capacitated Vehicle Routing Problems with Time Windows can be approximated through the use of distinct features which relate to the specific constraints of the problem. From a methodology perspective, in this exploration a computationally efficient method to allocate costs and test different regression models so as to find the best fit is found. This is done by combining different methods both to solve efficiently CVRPTW with a metaheuristic solver and to allocate costs accurately with a structured random sampling method to approximate the Shapley value.

In the case of simple CVRP the measure of distance is found to achieve a high explanatory power over the cost allocated to customers when dealing with less capacitated instances and low distances. Additionally, the distribution, and therefore the uncertainty, of the allocated costs grow with increasing distances where demand and vehicles' capacity limit the problem increasingly.

As for the CVRPTW, the main characteristics of the well-known Solomon instances modelled, such as the clustering, capacity and time window length, allow for the observation of how these variables and limitations of the problem affect the cost of the logistics operation. These observations are confirmed by analysing the cost allocated to each customer individually. Customer clustering and relaxation of time windows is found to decrease costs. Moreover, the ratio between customer demand and vehicle capacity and the amount of vehicles needed in general lead to increased costs. On the whole, it is found that costs are strongly correlated to potential savings from collaboration.

In order to approximate the calculated costs through a linear function, this study compiles different features to condense the variables affecting the problem. Consequently, through model fitting and feature selection, the most important features are defined and a cost prediction model is created. Later, the differences in models when working with different instances are assessed. Most importantly, it is concluded that the specifics of the CVRPTW, and the extent to which the constraints limit them, are decisive when approximating the cost allocation methods through a formula. Nonetheless, from the analysis of the instances in this study it is inferred that a high degree of accuracy can be achieved on the regression of costs allocated with only a small amount of features.

All in all, the classification of practical problems according to their constraints can result in great value to determine which factors are of utmost relevance. In the case of clustered customers, the relative distance to the depot and the measure of clustering are of strong importance. In contrast, with non-clustered customers, time windows gain major importance due to customers not being in proximity to each other. Further, different kinds of zoning suit better each kind of problem. Radial zoning adjusts better for non-clustered customers, whereas local zoning does so for clustered ones. With respect to time restrictions, it is observed that short scheduling horizons result in a higher variance of the costs allocated with a dependency on time window length. In conclusion, cost-allocating mechanisms and resulting vectors benefit from considering the different types of categorizations analyzed throughout our work; e.g. when collaborating, carriers can align on the clustering of neighborhoods and therefore choose to participate in only selected ones.

Two main pricing strategies are found, with their scope depending on the problem constraints: With or without a base price. It is derived that the classification of customer configurations among the categories used results in a better understanding on which strategy to utilize.

Further, this work assesses the logistic problem in the context of collaboration. The gains obtained by carriers when acting in a joint operation are determined and the impact of constraints on their costs and potential savings analyzed. The total sum of gains obtained by carriers when acting jointly is found to be positive in all instances studied. In general, there is a great degree of correlation found between the cost of carriers when acting in a stand-alone and the gains from collaborating. Nonetheless, a small fraction of customers is found where the grand coalition assigns a bigger cost than the stand-alone cost obtained with one carrier. An area of further study is identified where the underlying reasons for this result could be explored and whether or not the use of the Shapley value, lacking stability from a game theory application, could be improved as a cost allocation method. Furthermore, from a decentralized collaboration perspective, customers which can be served by other partners more effectively are found to be 56% more costly than the rest in each carrier's stand-alone operation signaling for the need of further developing this area of study.

This work is focused on deterministic scenarios. Potential paths of future research could be explored with stochastic demands or uncertainties in other variables. In contrast

with static assumptions in this research, dynamic changes with multiple periods could also be analyzed. Concerning the collaborative environment studied, auction scenarios represent a valuable area of further exploration. Moreover, the assumptions of having central planning and one main depot can be expanded into multi-depot and selective customer-sharing problems. To complement this study, the inclusion of more carriers and vertical integration could better enrich this research, with multi-modal logistics adding to the possibilities. In addition, the cost allocation mechanisms utilized could be applied in areas related to environmental impacts and their distribution. Finally, this work does not address the cost of sharing information and the extent to which it needs to be achieved for reasonable collaboration with regard to profits.

References

Agarwal, R., & Ergun, Ö. (2010). Network design and allocation mechanisms for carrier alliances in liner shipping. *Operations Research*, *58*(6), 1726–1742.

Akkerman, F., & Mes, M. (2022). Distance approximation to support customer selection in vehicle routing problems. *Annals of operations research*, 1–29.

Alfonso, V., Boar, C., Frost, J., Gambacorta, L., & Liu, J. (2021). E-commerce in the pandemic and beyond. *BIS Bulletin*, *36*(9), 1–9.

Anily, S., & Haviv, M. (2007). The cost allocation problem for the first order interaction joint replenishment model. *Operations Research*, *55*(2), 292–302.

Audy, J.-F., D'Amours, S., & Rönnqvist, M. (2012). An empirical study on coalition formation and cost/savings allocation. *International Journal of Production Economics*, *136*(1), 13–27.

Audy, J.-F., D'Amours, S., & Rousseau, L.-M. (2011). Cost allocation in the establishment of a collaborative transportation agreement—an application in the furniture industry. *Journal of the Operational Research Society*, *62*(6), 960–970.

Ballot, E., & Fontane, F. (2010). Reducing transportation CO2 emissions through pooling of supply networks: perspectives from a case study in French retail chains. *Production Planning & Control*, *21*(6), 640–650.

Berbeglia, G., Cordeau, J.-F., Gribkovskaia, I., & Laporte, G. (2007). Static pickup and delivery problems: a classification scheme and survey. *TOP*, *15*, 1–31.

Berger, S., & Bierwirth, C. (2010). Solutions to the request reassignment problem in collaborative carrier networks. *Transportation Research Part E: Logistics and Transportation Review*, *46*(5), 627–638.

Biermasz, J. (2012). *Report on the legal framework for horizontal collaboration in the supply chain* (tech. rep.).

Bratton, T. W., Mentzer, J. T., Foggin, J. H., Quinn, F. J., & Golicic, S. L. (2000). Supply chain collaboration: the enablers, impediments, and benefits. *Council of Logistics Management Fall MeetingCouncil of Logistics Management*.

Bräysy, O., & Gendreau, M. (2005). Vehicle routing problem with time windows, Part II: Metaheuristics. *Transportation science*, *39*(1), 119–139.

Bretzke, W.-R. (2013). Global urbanization: a major challenge for logistics. *Logistics Research*, *6*(2-3), 57–62.

Brown, R. (2016). Fundamentals of Correlation and Regression.

Caprara, A., & Letchford, A. N. (2010). New techniques for cost sharing in combinatorial optimization games. *Mathematical Programming*, *124*, 93–118.

Castro, J., Gómez, D., & Tejada, J. (2009). Polynomial calculation of the Shapley value based on sampling. *Computers & Operations Research*, *36*(5), 1726–1730.

Cheng, Y., Tan, J., & Lin, W.-H. (2013). Allocating the cost of empty railcar miles in a dynamic railroad network. *International Journal of Shipping and Transport Logistics*, *5*(3), 350–366.

Chinh, N. Q., Kim, H. C., Siwei, J., & NengSheng, Z. (2016). Collaborative vehicle routing problem for urban last-mile logistics. *2016 IEEE international conference on systems, man, and cybernetics (SMC)*, 001531–001536.

Christofides, N. (1979). The vehicle routing problem. *Combinatorial optimization*.

Clarke, G., & Wright, J. W. (1964). Scheduling of vehicles from a central depot to a number of delivery points. *Operations Research*, *12*(4), 568–581.

Cochran, W. G. (1977). *Sampling techniques*.

Cordeau, J.-F., Gendreau, M., & Laporte, G. (1997). A tabu search heuristic for periodic and multi-depot vehicle routing problems. *Networks: An International Journal*, *30*(2), 105–119.

Cordeau, J.-F., Laporte, G., Savelsbergh, M. W., & Vigo, D. (2007). Vehicle routing. *Handbooks in operations research and management science*, *14*, 367–428.

Cornuejols, G., & Harche, F. (1993). Polyhedral study of the capacitated vehicle routing problem. *Mathematical Programming*, *60*(1-3), 21–52.

Cruijssen, F., Borm, P., Fleuren, H., & Hamers, H. (2010). Supplier-initiated outsourcing: A methodology to exploit synergy in transportation. *European Journal of Operational Research*, *207*(2), 763–774.

Cruijssen, F., Bräysy, O., Dullaert, W., Fleuren, H., & Salomon, M. (2007). Joint route planning under varying market conditions. *International Journal of Physical Distribution & Logistics Management*, *37*(4), 287–304.

Cruijssen, F., Cools, M., & Dullaert, W. (2007). Horizontal cooperation in logistics: opportunities and impediments. *Transportation Research Part E: Logistics and Transportation Review*, *43*(2), 129–142.

Cruijssen, F., & Cruijssen, F. (2020). Applications of Cross-Chain Collaboration. *Cross-Chain Collaboration in Logistics: Looking Back and Ahead*, 91–105.

Cruijssen, F., Dullaert, W., & Fleuren, H. (2007). Horizontal cooperation in transport and logistics: a literature review. *Transportation journal*, *46*(3), 22–39.

Dahl, S., & Derigs, U. (2011). Cooperative planning in express carrier networks—An empirical study on the effectiveness of a real-time Decision Support System. *Decision Support Systems*, *51*(3), 620–626.

Dahle, L., Andersson, H., Christiansen, M., & Speranza, M. G. (2019). The pickup and delivery problem with time windows and occasional drivers. *Computers & Operations Research*, *109*, 122–133.

Dai, B., & Chen, H. (2011). A multi-agent and auction-based framework and approach for carrier collaboration. *Logistics Research*, *3*, 101–120.

Dai, B., & Chen, H. (2012). Profit allocation mechanisms for carrier collaboration in pickup and delivery service. *Computers & Industrial Engineering*, *62*(2), 633–643.

Dai, B., & Chen, H. (2015). Proportional egalitarian core solution for profit allocation games with an application to collaborative transportation planning. *European Journal of Industrial Engineering*, *9*(1), 53–76.

Dantzig, G. B., & Ramser, J. H. (1959). The truck dispatching problem. *Management science*, *6*(1), 80–91.

Defryn, C., Sörensen, K., & Cornelissens, T. (2016). The selective vehicle routing problem in a collaborative environment. *European Journal of Operational Research*, *250*(2), 400–411.

Defryn, C., Vanovermeire, C., & Sörensen, K. (2016). Gain sharing in horizontal logistic co-operation: a case study in the fresh fruit and vegetables sector. *Sustainable Logistics and Supply Chains: Innovations and Integral Approaches*, 75–89.

Deidda, D., Andreu, J., Perez, M., Sechi, G., Zucca, R., & Zuddas, P. (2009). A cooperative game theory approach to water pricing in a complex water resource system. *18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation*, 3252–3258.

Deng, X., & Papadimitriou, C. H. (1994). On the complexity of cooperative solution concepts. *Mathematics of operations research*, *19*(2), 257–266.

Derks, J., & Kuipers, J. (1997). On the core of routing games. *International Journal of Game Theory*, *26*, 193–205.

Desrochers, M., & Laporte, G. (1991). Improvements and extensions to the Miller-Tucker-Zemlin subtour elimination constraints. *Operations Research Letters*, *10*(1), 27–36.

Dowdy, S., Wearden, S., & Chilko, D. (2011). *Statistics for research*. John Wiley & Sons.

Dror, M. (1990). Cost allocation: the traveling salesman, binpacking, and the knapsack. *Applied Mathematics and Computation*, *35*(2), 191–207.

Engevall, S., Göthe-Lundgren, M., & Värbrand, P. (1998). The traveling salesman game: An application ofcost allocation in a gas and oil company. *Annals of Operations Research*, *82*(0), 203–218.

Engevall, S., Göthe-Lundgren, M., & Värbrand, P. (2004). The heterogeneous vehicle-routing game. *Transportation Science*, *38*(1), 71–85.

Ergun, O., Kuyzu, G., & Savelsbergh, M. (2007). Reducing truckload transportation costs through collaboration. *Transportation science*, *41*(2), 206–221.

Estévez-Fernández, A., Borm, P., Meertens, M., & Reijnierse, H. (2009). On the core of routing games with revenues. *International Journal of Game Theory*, *38*, 291–304.

Faigle, U., Fekete, S. P., Hochstättler, W., & Kern, W. (1998). On approximately fair cost allocation in Euclidean TSP games. *OR Spektrum*, *20*, 29–37.

Faigle, U., & Kern, W. (1992). The Shapley value for cooperative games under precedence constraints. *International Journal of Game Theory*, *21*, 249–266.

Fang, X., & Cho, S.-H. (2014). Stability and endogenous formation of inventory transshipment networks. *Operations Research*, *62*(6), 1316–1334.

Fernández, E., Roca-Riu, M., & Speranza, M. G. (2018). The shared customer collaboration vehicle routing problem. *European journal of operational research*, *265*(3), 1078–1093.

Ferrell, W., Ellis, K., Kaminsky, P., & Rainwater, C. (2020). Horizontal collaboration: opportunities for improved logistics planning. *International Journal of Production Research*, *58*(14), 4267–4284.

Fiestras-Janeiro, M. G., García-Jurado, I., Meca, A., & Mosquera, M. A. (2013). A new cost allocation rule for inventory transportation systems. *Operations Research Letters*, *41*(5), 449–453.

Fiestras-Janeiro, M. G., García-Jurado, I., Meca, A., & Mosquera, M. (2012). Cost allocation in inventory transportation systems. *TOP*, *20*, 397–410.

Fishburn, P., & Pollak, H. (1983). Fixed-route cost allocation. *The American Mathematical Monthly*, *90*(6), 366–378.

Flisberg, P., Frisk, M., Rönnqvist, M., & Guajardo, M. (2015). Potential savings and cost allocations for forest fuel transportation in Sweden: A country-wide study. *Energy*, *85*, 353–365.

Frisk, M., Göthe-Lundgren, M., Jörnsten, K., & Rönnqvist, M. (2010). Cost allocation in collaborative forest transportation. *European Journal of Operational Research*, *205*(2), 448–458.

Gansterer, M., & Hartl, R. F. (2018a). Centralized bundle generation in auction-based collaborative transportation. *Or Spectrum*, *40*(3), 613–635.

Gansterer, M., & Hartl, R. F. (2018b). Collaborative vehicle routing: A survey. *European Journal of Operational Research*, *268*(1), 1–12.

Gansterer, M., & Hartl, R. F. (2020). Shared resources in collaborative vehicle routing. *TOP*, *28*, 1–20.

Gansterer, M., Hartl, R. F., & Vetschera, R. (2019). The cost of incentive compatibility in auction-based mechanisms for carrier collaboration. *Networks*, *73*(4), 490–514.

Goel, A., & Gruhn, V. (2008). A general vehicle routing problem. *European Journal of Operational Research*, *191*(3), 650–660.

Göthe-Lundgren, M., Jörnsten, K., & Värbrand, P. (1996). On the nucleolus of the basic vehicle routing game. *Mathematical Programming*, *72*, 83–100.

Guajardo, M., Jörnsten, K., & Rönnqvist, M. (2016). Constructive and blocking power in collaborative transportation. *OR spectrum*, *38*, 25–50.

Guajardo, M., & Rönnqvist, M. (2015). Operations research models for coalition structure in collaborative logistics. *European Journal of Operational Research*, *240*(1), 147–159.

Guajardo, M., & Rönnqvist, M. (2016). A review on cost allocation methods in collaborative transportation. *International transactions in operational research*, *23*(3), 371–392.

Hamers, H., Borm, P., van de Leensel, R., & Tijs, S. (1999). Cost allocation in the Chinese postman problem. *European Journal of Operational Research*, *118*(1), 153–163.

Hansen, M. H., & Yu, B. (2001). Model selection and the principle of minimum description length. *Journal of the american statistical association*, 746–774.

Hezarkhani, B., Slikker, M., & Van Woensel, T. (2016). A competitive solution for cooperative truckload delivery. *Or Spectrum*, *38*, 51–80.

Hu, W., Dong, J., Hwang, B.-g., Ren, R., & Chen, Z. (2019). A scientometrics review on city logistics literature: Research trends, advanced theory and practice. *Sustainability*, *11*(10), 2724.

Iannaccone, G., Marcucci, E., & Gatta, V. (2021). What young e-consumers want? Forecasting parcel lockers choice in Rome. *Logistics*, *5*(3), 57.

Irnich, S., Toth, P., & Vigo, D. (2014). Chapter 1: The family of vehicle routing problems. In *Vehicle Routing: Problems, Methods, and Applications, Second Edition* (pp. 1–33). SIAM.

Jílková, P., & Králová, P. (2021). Digital consumer behaviour and ecommerce trends during the COVID-19 crisis. *International Advances in Economic Research*, *27*(1), 83–85.

Kallehauge, B., Larsen, J., Madsen, O. B., & Solomon, M. M. (2005). *Vehicle routing problem with time windows*. Springer.

Karaenke, P., Bichler, M., & Minner, S. (2019). Coordination is hard: Electronic auction mechanisms for increased efficiency in transportation logistics. *Management Science*, *65*(12), 5884–5900.

Krajewska, M. A., & Kopfer, H. (2006). Collaborating freight forwarding enterprises: Request allocation and profit sharing. *OR spectrum*, *28*, 301–317.

Krajewska, M. A., Kopfer, H., Laporte, G., Ropke, S., & Zaccour, G. (2008). Horizontal cooperation among freight carriers: request allocation and profit sharing. *Journal of the Operational Research Society*, *59*, 1483–1491.

Kritikos, M. N., & Ioannou, G. (2010). The balanced cargo vehicle routing problem with time windows. *International Journal of Production Economics*, *123*(1), 42–51.

Laporte, G. (1992). The traveling salesman problem: An overview of exact and approximate algorithms. *European Journal of Operational Research*, *59*(2), 231–247.

Laporte, G. (2009). Fifty years of vehicle routing. *Transportation science*, *43*(4), 408–416.

Lehoux, N., d'Amours, S., Frein, Y., Langevin, A., & Penz, B. (2011). Collaboration for a two-echelon supply chain in the pulp and paper industry: the use of incentives to increase profit. *Journal of the Operational Research Society*, *62*(4), 581–592.

Leng, K., Ya, B., Linbo, J., Han-Chi, F., & Van Nieuwenhuyse, I. (2018). Research on agricultural supply chain system with double chain architecture based on blockchain technology. *Future Generation Computer Systems*, *86*(1), 641–649.

Littlechild, S. C., & Owen, G. (1973). A simple expression for the Shapley value in a special case. *Management Science*, *20*(3), 370–372.

Liu, P., Wu, Y., & Xu, N. (2010). Allocating collaborative profit in less-than-truckload carrier alliance. *Journal of Service science and Management*, *3*(01), 143.

Lozano, S., Moreno, P., Adenso-Díaz, B., & Algaba, E. (2013). Cooperative game theory approach to allocating benefits of horizontal cooperation. *European Journal of Operational Research*, *229*(2), 444–452.

Massol, O., & Tchung-Ming, S. (2010). Cooperation among liquefied natural gas suppliers: Is rationalization the sole objective? *Energy Economics*, *32*(4), 933–947.

Miller, C. E., Tucker, A. W., & Zemlin, R. A. (1960). Integer programming formulation of traveling salesman problems. *Journal of the ACM (JACM)*, *7*(4), 326–329.

Moretti, S., & Patrone, F. (2008). Transversality of the Shapley value. *TOP*, *16*, 1–41.

Moretti, S., Patrone, F., & Bonassi, S. (2007). The class of microarray games and the relevance index for genes. *TOP*, *15*, 256–280.

Moulin, H. (1991). *Axioms of cooperative decision making*. Cambridge university press.

Muñoz-Villamizar, A., Montoya-Torres, J. R., & Vega-Mejía, C. A. (2015). Non-collaborative versus collaborative last-mile delivery in urban systems with stochastic demands. *Procedia CIRP*, *30*, 263–268.

Nguyen, C., Dessouky, M., & Toriello, A. (2014). Consolidation strategies for the delivery of perishable products. *Transportation Research Part E: Logistics and Transportation Review*, *69*, 108–121.

Osborne, J. W., & Waters, E. (2002). Four assumptions of multiple regression that researchers should always test. *Practical assessment, research, and evaluation*, *8*(1), 2.

Özener, O. Ö. (2014). Developing a collaborative planning framework for sustainable transportation. *Mathematical Problems in Engineering*, *2014*.

Özener, O. Ö., & Ergun, Ö. (2008). Allocating costs in a collaborative transportation procurement network. *Transportation Science*, *42*(2), 146–165.

Özener, O. Ö., Ergun, Ö., & Savelsbergh, M. (2011). Lane-exchange mechanisms for truckload carrier collaboration. *Transportation Science*, *45*(1), 1–17.

Özener, O. Ö., Ergun, Ö., & Savelsbergh, M. (2013). Allocating cost of service to customers in inventory routing. *Operations Research*, *61*(1), 112–125.

Pérez-Bernabeu, E., Juan, A. A., Faulin, J., & Barrios, B. B. (2015). Horizontal cooperation in road transportation: a case illustrating savings in distances and greenhouse gas emissions. *International Transactions in Operational Research*, *22*(3), 585–606.

Potters, J. A., Curiel, I. J., & Tijs, S. H. (1992). Traveling salesman games. *Mathematical Programming*, *53*, 199–211.

Prins, C. (2004). A simple and effective evolutionary algorithm for the vehicle routing problem. *Computers & operations research*, *31*(12), 1985–2002.

Puka, R., Duda, J., & Karkula, M. (2019). Improving of Selected Routing Strategies for Order Picking. *International Scientific Conference Transport of the 21st Century*, 363–373.

Ranieri, L., Digiesi, S., Silvestri, B., & Roccotelli, M. (2018). A review of last mile logistics innovations in an externalities cost reduction vision. *Sustainability*, *10*(3), 782.

Rawlings, J. O., Pantula, S. G., & Dickey, D. A. (1998). *Applied regression analysis: a research tool*. Springer.

Sanchez, M., Pradenas, L., Deschamps, J.-C., & Parada, V. (2016). Reducing the carbon footprint in a vehicle routing problem by pooling resources from different companies. *NETNOMICS: Economic Research and Electronic Networking*, *17*, 29–45.

Schmeidler, D. (1969). The nucleolus of a characteristic function game. *SIAM Journal on applied mathematics*, *17*(6), 1163–1170.

Schmelzer, H., Bütikofer, S., & Hollenstein, L. (2016). Kooperieren? ja! aber wie?: Chancen und Herausforderungen bei der Entwicklung einer Kooperationsplattform für urbane Güterlogistik in der Stadt Zürich. *Logistics Innovation*, *2016*(1), 16–19.

Schulte, F., Lalla-Ruiz, E., González-Ramírez, R. G., & Voß, S. (2017). Reducing port-related empty truck emissions: a mathematical approach for truck appointments with collaboration. *Transportation Research Part E: Logistics and Transportation Review*, *105*, 195–212.

Shapley, L. S. (1951). Notes on the n-person game. *Rand Corporation*.

Shapley, L. S. (1953). A value for n-person games. *Annals of Math. Studies*, *28*.

Shi, Y., Lin, N., Han, Q., Zhang, T., & Shen, W. (2020). A method for transportation planning and profit sharing in collaborative multi-carrier vehicle routing. *Mathematics*, *8*(10), 1788.

Solomon, M. M. (1987). Algorithms for the vehicle routing and scheduling problems with time window constraints. *Operations Research*, *35*(2), 254–265.

Song, L., Cherrett, T., McLeod, F., & Guan, W. (2009). Addressing the last mile problem: transport impacts of collection and delivery points. *Transportation research record*, *2097*(1), 9–18.

Sun, L., Rangarajan, A., Karwan, M. H., & Pinto, J. M. (2015). Transportation cost allocation on a fixed route. *Computers & Industrial Engineering*, *83*, 61–73.

Sykes, A. O. (1993). An introduction to regression analysis.

Tang, Q., Zhang, Z., Yuan, Z., & Li, Z. (2022). The Game Analysis of Information Sharing for Supply Chain Enterprises in the Blockchain. *Frontiers in Manufacturing Technology*, *2*, 880332.

Tarantilis, C. D. (2005). Solving the vehicle routing problem with adaptive memory programming methodology. *Computers & Operations Research*, *32*(9), 2309–2327.

Toriello, A., & Uhan, N. A. (2013). On traveling salesman games with asymmetric costs. *Operations Research*, *61*(6), 1429–1434.

Toth, P., & Vigo, D. (2014). *Vehicle routing: problems, methods, and applications*. SIAM.

van Campen, T., Hamers, H., Husslage, B., & Lindelauf, R. (2018). A new approximation method for the Shapley value applied to the WTC 9/11 terrorist attack. *Social Network Analysis and Mining*, *8*, 1–12.

Vanovermeire, C., & Sörensen, K. (2014a). Integration of the cost allocation in the optimization of collaborative bundling. *Transportation Research Part E: Logistics and Transportation Review*, *72*, 125–143.

Vanovermeire, C., & Sörensen, K. (2014b). Measuring and rewarding flexibility in collaborative distribution, including two-partner coalitions. *European Journal of Operational Research*, *239*(1), 157–165.

Vanovermeire, C., Sörensen, K., Van Breedam, A., Vannieuwenhuyse, B., & Verstrepen, S. (2014). Horizontal logistics collaboration: decreasing costs through flexibility and an adequate cost allocation strategy. *International Journal of Logistics Research and Applications*, *17*(4), 339–355.

VROOM Project. (2022). Vehicle Routing Open Source Optimization Machine. http://vroom-project.org/

Wang, Y., Ma, X., Xu, M., Wang, L., Wang, Y., & Liu, Y. (2015). A methodology to exploit profit allocation in logistics joint distribution network optimization. *Mathematical Problems in Engineering*, *2015*.

Wong, H., Van Oudheusden, D., & Cattrysse, D. (2007). Cost allocation in spare parts inventory pooling. *Transportation Research Part E: Logistics and Transportation Review*, *43*(4), 370–386.

Yang, F., Yang, M., Xia, Q., & Liang, L. (2016). Collaborative distribution between two logistics service providers. *International Transactions in Operational Research*, *23*(6), 1025–1050.

Yengin, D. (2012). Characterizing the Shapley value in fixed-route traveling salesman problems with appointments. *International Journal of Game Theory*, *41*, 271–299.

Yilmaz, O., & Savasaneril, S. (2012). Collaboration among small shippers in a transportation market. *European journal of operational research*, *218*(2), 408–415.

Yule, G. U., & Kendall, M. (1968). *An introduction to the theory of Statistics*. Charles Griffin; Co. Ltd.

Zakharov, V. V., & Shchegryaev, A. N. (2015). Stable cooperation in dynamic vehicle routing problems. *Automation and Remote Control*, *76*, 935–943.

Junior Management Science

www.jums.academy
ISSN: 2942-1861

Rewarding Creativity: The Moderating Role of Personality

Maryam Rassouli-Baghi

University of Amsterdam

Abstract

The aim of this paper is to find out if Openness to Experience and Extraversion have an interacting effect on the relationship between monetary rewards and originality. Therefore, in an online experiment which was set up in two parts - the first part measured participants personality level whereas the second part included a divergent thinking task, where the originality of ideas was measured - I find that Openness to Experience and Extraversion positively influence creativity. Further, it is assumed, that monetary rewards positively influence an individual's creativity, as those individuals being rewarded for creativity performed more creative. However, analysis indicates that there is no interaction effect as assumed. As the ability of creative thinking is essential for companies and can even lead to a competitive advantage, the findings of this study provide useful insights about the relevance of personality on creativity for theory and practice.

Keywords: creativity; divergent thinking, performance; personality; reward sensitivity

1. Introduction

Management problems often require new views and creative insights. Especially due to current change and global competition, creativity is an important issue for managers and companies, as it contributes to discovering new opportunities (Biraglia & Kadile, 2017) and innovation in organizations (Amabile, 1998). Oldham and Cummings (1996) conclude that enhancing the creative performance of employees is necessary for organizations to achieve competitive advantage. As a result, interest has increased in identifying the conditions which may influence employee's creativity (Madjar et al., 2002). Hence, the ability of creative thinking in management is essential, which is typically measured through creativity tests such as divergent thinking tasks (Scratchley & Hakstian, 2001).

Consistent with this, J. Zhao et al. (2020) measured creative performance and proposed the use of reward systems to ensure standards. The impact of rewards on creativity has received much attention in the literature, but only little agreement on the effect of rewards (Shalley et al., 2004). According to Amabile (1996), extrinsic motivation negatively influenced creativity, whereas Eisenberger (1992) found a positive influence by demonstrating the impact of rewards on personal competencies. Additionally, in a study conducted by Joussemet and Koestner (1999) the authors also found an increased creative performance by using rewards but noticed a decreased quality of answers. In contradiction, other studies (Yoon, Sung, Choi, et al., 2015) could not find any significant relationship between rewards and creativity, which demonstrates the ongoing ambiguous and unclear impact of rewards on creative performance.

Following up on these findings, there has also been a growing interest in the study of personality and its impact on creativity. For instance, Sung and Choi's (2009) analysis of the moderating effect of extrinsic motivation on the personality-creativity relation confirmed a positive influence of extrinsic motivation on openness to experience, which resulted in greater creative performance. Herrmann and Nadkarni (2014) found out that extraversion, openness and emotional stability could be related to CEO personality and creative performance. Similar results can be stated by Harrison et al. (2019) which also propose to take firm particular situations into account, which may influence the relationship between personality traits. Other studies focused on the relation among openness, divergent thinking, and creativity in management (Scratchley & Hakstian, 2001) and found evi-

DOI: https://doi.org/10.5282/jums/v9i1pp1269-1285

dence for openness as a key personality and divergent thinking as a key cognitive ability for creativity. This was also confirmed by Myszkowski et al. (2015) who claimed that personality predicts creativity and that openness to experience can be related to the divergent-exploratory part of the creative process. In contradiction to these results, Puryear et al. (2017) could not find any direct influence of personality on creativity.

Although several publications have appeared in recent years documenting the impact of motivation, creativity and personality, research has proceeded only slowly in testing the effects in this relationship. Further, research in this area has resulted in almost as much contradiction as agreement. Additionally, relatively less attention has been paid to the possibility that personality characteristics might have an impact on the reward-creativity relationship. Therefore, research evidence suggests exploring the type of employees that are more likely to engage in creative behavior (Zhou & George, 2001) or how personality can lead to a specific behavior, which in turn relates to firm performance (H. Zhao et al., 2010). Further, as the effect of extrinsic motivation on creativity is contradictory, more research is needed in this area focusing on its impact on creative performance (Shalley et al., 2004) and especially on divergent thinking, in this study measured as originality.

In response to the scholarly call for more research in these areas, the purpose of this study is to contribute to the ongoing discussion by reviewing existing literature to extend our understanding of the impact of personality on the reward-creativity relationship in the strategy field. In other words, I want to test individuals' perception of rewards, and see if specific personality traits, namely openness to experience and extraversion have a higher or lower reward sensitivity and therefore, perform more or less creative. The research question guiding this study is therefore, *How does personality moderate the relationship between monetary rewards and originality?* To research this, I conducted an online experiment among individuals with a (strategic) management background and tested their levels of personality as well as their ability for divergent thinking. Interestingly, this study approved openness to experience and extraversion being relevant factors for creativity, as well as financial rewards having an impact on creative performance, even though the moderating effect of personality could not be found in this sampling.

This study contributes to existing literature by making a new contribution to the ongoing debate on the reward-creativity puzzle. Research testing the impact of rewards on performance have yielded inconsistent findings (Amabile, 1996; Eisenberger, 1992) and only a few studies tested its impact on divergent thinking (Sung & Choi, 2009). However, to the author's best knowledge, no publications can be found in the literature discussing the issue of personality as a moderating factor between extrinsic motivation and creativity. Hence, this is the first study that empirically tests the moderating role of personality on extrinsic rewards and creativity as originality. Useful insights for theoretical application can be gained, as there is no literature focusing on this research gap, although the scholarly call for research in this area is clearly defined. Further, this study has also practical implication for management. The results of this study can be useful for managers in terms of recruiting, appraisals, promotions or motivating employees, as it provides insights into the impact of personality characteristics on creativity.

The section that follows presents the literature review and hypotheses underlying this study. Next, and drawing upon data from 141 participants operating in the management field, the methodology and empirical findings are presented. The study concludes with a discussion of the results, their theoretical and managerial implications, future research suggestions, and a conclusion.

2. Literature Review and Hypotheses

2.1. Creativity and the Importance of Divergent Thinking

This study adopts the dominant scholarly definition of creativity using a product-based approach to creativity. Creativity is defined as the production of novel and useful ideas, solutions, or products (Amabile, 1998) and is essential in a volatile business climate, where firms compete to create or maintain their competitive advantage (Walker & Jackson, 2004). Therefore, great effort has been devoted to the study of creativity (Byron & Khazanchi, 2012; Dollinger et al., 2004; George & Zhou, 2001; Mumford, 2000; Runco, 2007).

According to Barron and Harrington (1981), one way to measure creativity is through divergent thinking tests, which are often used in the literature (Chamorro-Premuzic & Reichenbacher, 2008; McCrae, 1987; Myszkowski et al., 2015). Additionally, the authors propose two factors concerning divergent thinking, namely the field-specific relevance of divergent thinking abilities and the state of knowledge regarding creativity in a given field, and the role of intelligence in divergent thinking. In regards to test instructions and test conditions, the authors claim that when divergent thinking instructions to be creative were combined with scoring procedures, the results were more significant regarding creativity. They further propose that the divergent thinking ability and traditional measures of intelligence vary widely depending on the divergent thinking test, the sample heterogeneity, and the testing conditions.

However, research has identified contradictory views on differencing creativity and divergent thinking. While some authors use both concepts similarly (Silvia et al., 2008), others argue for its need to differentiate, as the former may describe several abilities of acting creative, while the latter describes a cognitive process for creative problem-solving and idea generation (Runco et al., 2008). Nonetheless, it has been noted that most of the literature uses both terms consistently. In fact, even though over the last several decades there has been a discussion regarding distinctions of concepts, this study could not find any contrasting views or research denying the importance of divergent thinking.

It was Guilford (1956) who first proposed a difference between the two information processing modes divergent

thinking and convergent thinking, which can be understood as creative problem solving. Managers often not only have to diverge and find solutions but also converge and decide on the best solution (Myszkowski et al., 2015). Hence, while divergent thinking involves the search for a solution to a given problem for which there are multiple possible answers (Runco, 2007), convergent thinking can be understood as finding and selecting the one best response (Lubart, 2016). Additionally, the ability of divergent thinking results in an advantage to generate and construct a wide range of novel and original ideas, which builds the basis for creativity (McAuliffe, 2016).

To measure creativity, divergent thinking tasks have received much more attention in the literature and are often used as they provide the most reasonable results (Runco, 2007). Tasks measuring divergent thinking consist of different open-ended questions asking for creative uses and solutions for everyday objects. The most common scoring methods used in the literature are based on originality, fluency, and flexibility. Originality represents the number of unusual or unique ideas, fluency stands for the number of ideas in general, and flexibility refers to the number of different categories implied by the ideas (Runco & Acar, 2012). The scoring method in this study will focus on originality.

In a study testing how middle and senior managers' characteristics influence decisions, Behrens et al. (2014) for instance found out that the ability of divergent thinking increases with the level of experience. In contrast to this, Ames and Runco (2005) tested the divergent thinking ability of entrepreneurs and found lower scores regarding their originality. These results could indicate that entrepreneurs rely too much on their experience resulting from their own business knowledge, instead of their skills of generating novel ideas. This may also imply the importance for managers, as their ability to creative thinking changes over time and hence their contribution to performance. Therefore, managers should not only focus on experience, but also on their skills for divergent thinking.

To conclude, it seems reasonable to reinforce the assumption that divergent thinking can be seen as a key cognitive ability for creativity (Scratchley & Hakstian, 2001) and hence, that originality is essential for strategy and management. Now that we have determined the importance of creativity, the question arises as to how divergent thinking can be positively influenced or even increased. A possible answer to this question will be addressed in the next chapters.

2.2. The Influence of Rewards on Divergent Thinking

Pay for performance is the most common practice used by organizations to increase employee performance and motivation (Gupta & Shaw, 2014). When employees are being rewarded for creativity, they understand they have to provide novel but also useful ideas (Sue-Chan & Hempel, 2016).

Thus, scholars have identified two different types of motivation, intrinsic and extrinsic motivation, whereby an individuals' perceived self-determination plays an essential role (Deci & Ryan, 1985). Theory indicates that intrinsic motivation can be understood as a desire to act based on enjoyment and interest (Amabile, 1998), while extrinsic motivation can be interpreted as a motivation on the result of action provided by others. In addition, Amabile (1998) claims that people are even most creative, when they feel intrinsically motivated, satisfied and challenged by the work itself. Malik et al. (2019) claim that employees with intrinsic task motivation spend their resources and time in creative efforts without the wish or expectation of getting rewarded. However, intrinsic and extrinsic motivation can appear in both, a temporary state form affected by the environment or a more stable personality trait form, relatively consistent across time and situations. Further, findings also suggest that intrinsic and extrinsic motivation can be seen as orthogonal, which indicates that individuals can be for instance motivated by both, money and personal challenge (Amabile, 1993).

Nonetheless, scholars argue for its need to differentiate the types of motivation when it comes to creativity and claim that "motivating creative performance is fundamentally different from motivating routine performance" (Byron & Khazanchi, 2012, p. 809). Further, there are also different research streams when analyzing the effect of rewards on creative performance. While some authors focus more on cognitive effects and argue that rewards undermine intrinsic motivation as individuals perceive rewards as controlling which decreases creativity (e.g. Joussemet and Koestner, 1999), others take its' behavioral effect into account and claim that rewards provide information which may guide an individuals' goal-directed behavior which in turn results in increased creativity (e.g. Eisenberger, 1992).

When testing the effect of rewards, scholars predict that rewards implying regulation or control may decrease an individuals' perceived competence, intrinsic motivation and creative performance. A meta-analysis conducted by Byron and Khazanchi (2012) differentiates between (1) creativity-contingent rewards, which are rewards given for creativity, (2) performance-contingent rewards, which are rewards given for performance without defining a creativity criterion, and (3) completion-contingent rewards, which are rewards for completing tasks without stressing creativity as a factor for valuation. According to their results, the former seems to be the one most positively related to creativity, as it clearly directs an individuals' effort toward creativity and hence, increases creative performance.

The authors have further identified several moderators that affect the relationship between rewards and creativity, namely clear definitions for creativity criterions, positive feedback on performance, and balanced autonomy. This indicates that not only the nature of rewards is essential (Malik et al., 2015), but also the context in which the reward might be offered (Amabile & Gryskiewicz, 1989; Amabile et al., 2002; Hackman & Oldham, 1980; Oldham & Cummings, 1996) and the motivation behind (Amabile, 1998).

Moving on to extrinsic motivation, the situation becomes more complex. Despite the popularity of using extrinsic motivators such as monetary rewards to enhance creativity, there

is still little agreement on the effect on creative performance (Shalley et al., 2004). According to Amabile (1996) and Deci et al. (2001), rewards undermine employee creativity. As already indicated, research proposes that individuals may perceive rewards, namely performance-contingent rewards, as negative because they feel their performance is being determined based on controlling reward mechanisms (Malik et al., 2015). Perry-Smith and Mannucci (2017) even argue that rewards might negatively affect the generation of novel ideas because they might detract from the cognitive thinking process.

To test this, Kruglanski et al. (1971) asked students in an experiment to think of adequate titles for a paragraph; no further information was given about the nature of the titles. The authors noted that those promised a reward produced fewer creative answers, than others not receiving a reward. In response to this study, Eisenberger and Rhoades (2001) used the same test and asked participants to generate creative titles for a story. The main difference this time was that participants were informed that the experiment was about rewarding creative performance. Interestingly, in this study, those participants being rewarded for creativity, achieved much better results. In fact, this leads to the assumption that participants might perform better in creative tasks, as long as it is clearly defined, that the reward is given for creativity. In addition to this, research proposes that rewards only enhance creativity, when they are perceived as important (Yoon, Sung, & Choi, 2015).

But there are also other studies following different approaches and proposing other views. Huo (2020) for instance, tested the effect of rewards on divergent thinking by using three different incentive schemes. The first condition was a fixed incentive regardless of the performance, for the second condition the incentive was an average amount plus additional money for each solution and the last condition provided a conceptual replication of condition one including public recognition. As a result, even though the creative performance was higher for the second condition, no significant correlation could be found between performance incentives and divergent thinking. Also, other studies (George & Zhou, 2002) tested the effect of rewarding creativity on divergent thinking and could not identify any significant correlation.

In addition to the ongoing debate, Bonner et al. (2000) propose that the more complex the cognitive task, the less likely incentives may lead to improved creativity. Baer et al. (2003), tested the relationship between extrinsic rewards and creativity among employees and found a positive relationship between extrinsic rewards and creativity for simple tasks and a weak relation for more complex tasks. The results obtained by Webb et al. (2013) are in line with these propositions and add that assigning participants to easy tasks with a fixed reward leads to higher creativity, than those assigned to complex and challenging tasks or target-based pay.

In contradiction, F. Li et al. (2017) report higher perceptions of rewards, the more challenging the task. This finding is also in accordance with the research of Byron and Khazanchi (2012) who claim that especially for rewards given for creative performance, task complexity plays an essential role. The authors argue that a higher task complexity is more likely to increase an individuals' perceived competence and will also be more intrinsically motivating, which is why individuals will be more engaged and creative. In this context it should also be considered that employees might become more creative when they start feeling more confident in their work tasks, which could be a possible solution for these contradictory findings (Tierney & Farmer, 2002).

While some studies reported a negative influence and other studies found no influence, Eisenberger (1992) in contrast found a positive influence on creative performance by stating that rewards have an impact on personal competencies. These results can also be approved by Joussemet and Koestner (1999) which also noted an increased performance by rewarding creativity, but a decreased quality of answers. In contradiction, Byron and Khazanchi (2010) confirm the positive influence of creativity-contingent rewards on creativity itself, but state that extrinsic rewards may harm the overall performance.

Following the assumption that rewards may positively influence creative thinking, Kachelmeier et al. (2008) found evidence for the proposition that "quantity-based compensation improves quantity and that creativity-based compensation improves average creativity" (p. 343). Other studies (Eisenberger & Rhoades, 2001) proposed that specific contingency rewards for creativity may increase extrinsic motivation, while general rewards for high performance may rather increase intrinsic motivation. Additionally, it was confirmed that giving rewards for creativity for one task, also increased the creativity for subsequent tasks.

Moreover, Saether (2020) conducted a study among over 300 employees testing the impact of rewards on their divergent thinking. The author assigned the participants randomly into five groups, one of which performance was not rewarded. The other groups were manipulated in terms of the amount of money and instructions that should influence their motivation and perception of justice. As a result, creativity was increased when rewards were evaluated as fair. This also shows the importance of how the perception of fairness influences creative performance among participants, which can also be translated to employees in general. However, since Saether (2020) introduces another variable, namely the time at which the reward is to be paid, I argue that it might be no longer clear whether the motivation to be creative is related to the expected reward or whether the time has a stronger influence. It may be possible that some participants would have provided more creative results, but given different timelines, measuring the impact of the reward itself becomes ambiguous.

In fact, there are different views and results on the effect of rewards on performance or creativity, all stating different reasons for why or why not there might be a positive or negative outcome when using rewards. Kohn (1993) for instance claims that when organizations make use of rewards and incentives, people become even less interested in their work and wait for incentives before expending effort.

To conclude, even though research in this area testing the effect of monetary rewards on divergent thinking is contradictory, recent literature seems to provide more arguments for a positive influence on creativity, as long as it is clearly stated what the reward is given for. In this sense, I believe that this study will provide similar positive results. Further, this study not only wants to understand the reasons for these contradictory findings but also adds to existing literature by offering a new and different view on the rewards-creativity puzzle, which hasn't been solved so far. In fact, I assume that based on the findings the relationship between rewards and creativity might be more complex, which leads to the assumption that other moderators are at play here and might possibly impact the perception of rewards or the ability to creative thinking and problem solving.

Therefore, this study wants to explain those differences by taking personality traits into account, to test their influence. On the one hand I argue that personality might predict the creativity of individuals' and on the other hand there might be differences in terms of how rewards are perceived and how this might impact an individuals' performance based on his or her personality. Even though literature has mentioned their importance, no research has clarified their relationship. Therefore, in a first step, this study assumes a positive impact of rewards on creative performance.

Hypothesis 1. Monetary rewards for creativity are positively related to originality.

2.3. The Influence of Personality on Divergent Thinking

The study of personality is a very large field with an unmanageable number of theories, models, or measurements and there has been research in many areas testing if personality characteristics can be associated with creative achievements. Interestingly, there is a stable set of core characteristics such as broad interests, attraction to complexity or intuition pointing out how a creative personality could look like (Barron & Harrington, 1981).

In contradiction, not all scholars agree with the assumption of linking personality characteristics to creative performance. Zhou (1998) criticizes the dominant view of existing research on judging ideas based on their creativity and claims that mental processes through which those ideas are developed should be stressed more in research. Additionally, the author states that contextual variables may be more effective in predicting creativity than personality.

However, Amabile (1998) states that creative thinking refers to an individual's ability to solve problems, which also depends on his or her personality. Therefore, especially the personality of CEOs, managers' and leaders has emerged as a relevant topic in the strategic management field when examining personality traits, as their personalities can have a major influence on strategic decisions and may determine performance implications (Peterson et al., 2003). According to this, some personality traits might predict a stronger focus on tasks, flexibility or risk-taking, while other traits might enhance passivity or dominance (Herrmann & Nadkarni, 2014). Amabile (1983) claims that besides other factors, personality characteristics such as self-discipline, perseverance or independence are creativity-relevant skills and hence, related to creative thinking. In addition, Mumford (2000) predicts that creative individuals display a high degree of autonomy and are typically the ones developing new ideas.

Considering that the Five Factor Model (FFM) is one of the most known and used models to measure personality, scholars have already demonstrated its generalizability, validity and reliability in various studies (Costa & McCrae, 1992a). Further, other studies (H. Zhao & Seibert, 2006) have shown that the FFM traits are important drivers of individual human behavior and performance. Hence, this model, more specifically a shorter version of it (BFI; John and Srivastava, 1999), is also used for this study to measure its relation to creativity.

As Openness to Experience and Extraversion are generally predicted to have positive relationships to creative performance, this study wants to test if these results can also be achieved for divergent thinking, namely originality. Therefore, the focus lies on these two traits when testing personality and the creativity rating is based on the number of unusual or unique ideas (Runco, 2007).

2.3.1. Openness to Experience

Leaders with a high level of openness are more likely to express new ideas, are more flexible and open to changes and new experiences (Judge & Bono, 2000). Further, they seek out information, are able to identify more creative solutions for problems and are better in sensing and seizing opportunities (Shane et al., 2010).

Openness to experience "seems to be the most strongly tied to creativity" (Runco, 2007, p. 296) and is predicted as being a key personality for creativity (Dai et al., 2019; Harrison et al., 2019; Herrmann & Nadkarni, 2014; Nadkarni & Herrmann, 2010; H. Zhao & Seibert, 2006; H. Zhao et al., 2010). This is also confirmed by other studies testing the relationship between creativity and personality (Dollinger et al., 2004).

Myszkowski et al. (2015) for instance, found that managers with high levels of openness were able to find more ideas on management problems in divergent thinking tasks than others. To test this, they asked over 100 management students and designed their own creativity measures. The experiment consisted of two tasks each for divergent and convergent thinking and they rated creativity based on fluency, which represents the number of ideas. But also, other studies have consistently identified the positive relationship between openness to experience and divergent thinking and even claim that "creativity is uniquely related to openness to experience" (McCrae, 1987, p. 1263).

Furthermore, King et al. (1996) tested the relation among personality, creative ability and creative accomplishments among psychology students. The authors used the verbal component of the Torrance Test of Creative Thinking (TTCT) which consists of six different tasks such as unusual uses, or

supposing tasks. All answers were rated on two criteria, fluency and originality and as expected, openness to experience is positively related to both criteria. What the authors interestingly point out in the discussion section is the question why individuals with high levels of openness to experience give more creative answers. By trying to find an answer to this question they predict that those individuals might value creativity or novelty differently and might recognize potential situations as more creative than others.

Even from another point of view, namely a neuroscience view, W. Li et al. (2015) find support for the assumption that openness might positively influence the ability of divergent thinking. In fact, the authors propose that individuals with personality characteristics such as high levels of imagination, curiosity, challenge or risk-taking are more likely to have scientific discoveries than others. They conclude with suggesting that openness plays an important role in shaping an individual's creativity.

However, besides openness being positively related to creativity, George and Zhou (2001) also found out that creative answers were highest, when those individuals had unclear ends on their jobs and unclear means. This implies that the more freedom and autonomy people have, and less structure is given in tasks, the higher the ability to develop their creativity.

In fact, all these results indicate that openness to experience is a necessary personality trait in order to explore a high number of creative solutions to problems. Interestingly no literature could be found stating the opposite or presenting negative results for this trait when measuring creativity. It even seems impossible to assume different results as those presented in this study, since openness is being present as a key factor for creativity.

To conclude, there is no reason to expect different results for originality in a more strategic context as I assume that individuals working strategically also rely on openness and will give creative answers.

Hypothesis 2a. Openness to experience will be positively related to originality.

2.3.2. Extraversion

Individuals with a high level of extraversion tend to be enthusiastic and ambitious, which results in a proactive behavior by actively engaging in tasks and trying to find novel ideas (Raja et al., 2004). In contradiction, a low level of extraversion indicates a more quiet, introverted behavior (Costa & McCrae, 1992a).

However, several authors have tested the effect of extraversion on creativity, some of them also included additional variables in their analysis to test the importance of personality. According to Sung and Choi (2009, p. 944), their study indicates that "extraverts tend to seek novel ways of doing tasks and (...) confront problems instead of avoiding them" which leads them to propose a higher creative performance for extraverts. Furthermore, Gocłowska et al. (2019) tested the relation between extraversion, divergent thinking and creating novel ideas and found a positive relationship. Further, the author proposes that extraversion is, besides openness to experience, one of the creativity-related personality traits and conclude with confirming that novelty seeking is linked to extraversion and leads to greater divergent thinking.

Following this assumption, Kaspi-Baruch (2019) for instance, tested the moderating effect of goal orientation on extraversion and creativity and proposed a positive relationship. The experiment was conducted online using an adapted version of the Alternative Uses Task (AUT; Guilford, 1967) to test creativity. One point was awarded for every original answer. Indeed, as the results indicate, extraverted individuals seem to be the most creative, when they are motivated and oriented toward learning. A possible limitation I see in this study is that it might not be very representative, since 80% of the participants are female and the answers could therefore be biased.

In their study, Chamorro-Premuzic and Reichenbacher (2008) explored the effects of the Big Five personality traits and threat of evaluation on divergent and convergent thinking among students. Their results show that extraversion indeed predicts divergent thinking and is an important trait for creativity. This is explained by arguing that extraverted individuals might have an intrinsic motivated advantage in divergent thinking tasks and outperform their introverted counterparts, especially when they are being evaluated. However, the authors state that extraversion can be seen as facilitator for divergent thinking and is also related to actual creativity and not only to measured creativity.

Nonetheless, Nadkarni and Herrmann (2010) also confirm a positive relationship between extraversion and creative thinking. In their study, the authors tested the relationship between CEO personality, strategic flexibility, and performance. They state that strategic flexibility influences creativity and hence, tested the moderating effect of it on extraversion using the NEO Five Factor Inventory. Moreover, Weiss et al. (2020) noted a strong correlation between extraversion and divergent thinking and pointed out that extraversion was the strongest trait measured in their study related to divergent thinking.

Indeed, extraversion is consistently determined to have a positive relationship with creativity, regardless of whether this involves the personality of CEO's, managers or students. It seems likely to assume that according to the incorporated literature, extraverted individuals are more likely to generate more ideas in creativity tasks and will engage proactively to perform better than others. In fact, it is also quite likely that we will find the same result for originality in this study, since this is a part of divergent thinking and no contradictory results could be found stating the opposite.

Hypothesis 2b. Extraversion will be positively related to originality.

2.4. The Moderating Role of Personality

Given the contradictory findings of the effect of rewards on divergent thinking, research should explore the reasons under which rewards have positive, negative or neutral effects on divergent thinking (Baer et al., 2003) by taking other views into account. As part of a possible solution to this puzzle, I focus on personality in this section and want to test the reward sensitivity of individuals' personality traits, namely openness to experience and extraversion by offering a reward for creativity. With that in mind, it should be considered that individuals' personalities have different characteristics which respond differently to motivational factors such as extrinsic rewards (Shalley et al., 2004; Shaw et al., 2003).

It might be reasonable to argue, that some personalities may have a higher sensibility for rewards than others. In this context, reward sensitivity can be understood as "an incentive motivational state that facilitates and guides approach behavior to a goal" (Depue & Collins, 1999, p. 495). There are also other similar research approaches in this field testing the impact of personality on income and salaries. Hence, literature suggests that it is possible that personality may act as a moderator by interacting with rewards and how they are perceived (Wu & Zumbo, 2008).

Despite extensive research, no literature was found testing this moderating effect in this relationship. Most research in this area is focused on testing the rewarding effect on creative performance by taking different variables into account, such as fairness or different forms of manipulations (Saether, 2020), to contribute to the contradictory reward-creativity puzzle. Other studies tested the effect of all five personality traits on creativity to discover specific traits related to divergent thinking (King et al., 1996) or focused on differences between intrinsic and extrinsic motivation regarding creativity (Malik et al., 2019).

Therefore, related literature and results on openness to experience, extraversion, financial incentives, rewards and creative performance will be drawn together to present a status-quo in this field.

2.4.1. Openness to Experience

Openness to experience involves interest in novel things, ideas or knowledge and is related to characteristics such as being imaginative or adventurous (Shi et al., 2016). Furthermore, as we have already stated, this personality trait is the closest one related to creativity (Kaufman et al., 2015; Nadkarni & Herrmann, 2010). Hence, even though its relation to creativity might be clear, testing if there is a possible interaction effect on the reward-originality relationship might reveal new insights which could contribute to this debate.

An interesting study has been conducted by Sung and Choi (2009) who tested the moderating effect of intrinsic and extrinsic motivation on personality and creativity among business school students. Even though the moderator is a different one compared to my study, we find interesting implications here. Besides their proposition that extrinsic motivation might have a positive effect on creativity, the authors add that extrinsic motivation may offer a stage where individuals can behave in accordance with their personality trait. According to their study design, they argue that motivation might create a setting in which an individual's openness to experience can be activated to enhance creative performance. Indeed, the authors' hypothesis was supported, confirming a positive interaction effect of extrinsic motivation on openness, resulting in a greater creativity. However, what should be viewed critically is the extrinsic motivator the authors use. Most research provides monetary rewards to measure the impact of extrinsic motivation, while the experiment in this study provided gift certificates to students. In my opinion, I think it is necessary here to critically question the influence of the extrinsic motivator. Moreover, the authors talk about creative performance, but without going into more detail about what exactly is meant by this.

In another study (Heineck, 2011), the author tested the relationship between personality traits and market success in the United Kingdom by taking monetary effects into account. It should be noted that although this study does not measure creativity, it still focuses on performance differences as an outcome. The author used data from the British Household Panel Survey and a fifteen-item questionnaire to capture the personality dimension. The results indicate a positive relationship between openness to experience and wages, which may also be considered for rewards.

This raises the question if besides the positive relationship between openness and wages, there is also evidence for openness reacting more sensitive than other traits to rewards and payments. To test this, Vandenberghe et al. (2008) conducted a study among 967 participants to find out, which personality trait was most attracted to total rewards and its components such as variable or indirect pay, but also prestige and the quality of social relationships at work. Their results showed that openness best predicts the importance employees give to total rewards. Regarding bonus payments, the authors claim that "people who are curious and open to experience tend to ascribe greater prominence to bonuses (...)" (p. 442).

Openness to experience has proven to be a trait that has a positive impact on creativity (Myszkowski et al., 2015; Runco, 2007). Besides, Sung and Choi (2009) have shown the interaction effect of extrinsic motivation on openness and creativity. Hence, if there is a positive interaction effect of extrinsic motivation on openness, it might be reasonable to assume similar results for openness interacting with rewards, by offering a similar stage for individuals' to behave more or less creative according to their reward sensitivity.

Nevertheless, there is also research proposing extraversion instead to be the personality trait most attracted to compensation (Lucas & Diener, 2001), as they argue that those individuals are more risk tolerant and engage more in social interactions. Contrary, openness to experience is linked to seeking new challenges, development and new opportunities (Barrick & Mount, 1991).

In fact, I assume that it seems likely that individuals with high levels of openness, as this trait stands for making new

experiences and being open for novel ideas, might perceive extrinsic rewards as positive which in turn might increase originality. Therefore, this study predicts a higher reward sensitivity for those individuals with high levels of openness.

Hypothesis 3a. Openness to experience will positively moderate the relationship between rewards and originality.

2.4.2. Extraversion

Extraverted individuals are open, sociable, talkative, have positive emotions and thoughts and enjoy interaction with others (Gocłowska et al., 2019; King et al., 1996). Research also indicates that extraverted individuals might have a stronger reaction to reward-approach situations than others (Smillie et al., 2012), but more focus is paid on brain systems and connections in the literature, rather than on taking a behavioral approach.

Interestingly, when it comes to defining the core features of extraversion there are contradictory findings in the literature, finding evidence for either social interaction or reward-sensitivity being the key determinants for extraversion (Ashton et al., 2002). The authors state that it might be possible that "extraverts are sociable because (a) reward sensitivity is the core of Extraversion and (b) social situations tend to be rewarding" (p. 245).

In a study conducted by Speed et al. (2018), the authors tested the relationship between extraversion on reward sensitivity. Their results indicate that extraversion was associated with reward sensitivity, but only when neuroticism was low. Interesting in this study was further the approach the authors used to test the participant's reward sensitivity. They presented two identical doors on a computer screen to the participants, asking them to select one. Additionally, the participants were told that they could either win or lose money, depending on which door they chose. Participants did not know where the reward was hidden. In my opinion, the fact that participants could also lose money is interesting, as no other studies could be found testing similar effects, but also raises the question if being rewarded and losing money can be tested equally. I wonder what specifically the study is attempting to establish with this. On the one hand, it agrees with previous findings and hypothesizes that extroverts react positively to rewards. But I wonder if the same is true for losing money. However, another critique on this study regarding its validity, is that the study has been conducted on adolescent girls only and therefore cannot be generalized.

In line with these findings the American Psychological Association conducted a study (2021) on approximately 300 participants, testing possible effects of personality traits such as extraversion but also neuroticism or depression on reward-sensitivity. From a neuroscience point of view, their findings suggest that extroverts may have certain connections in their brain system that tend to perceive reward systems as a greater incentive than others. However, even though the foundation for reasoning is a different one, the results stating a higher reward-sensitivity for extraverts are similar.

In returning to a behavioral approach, Gocłowska et al. (2019) focused more on the extraversion-creativity relationship and tested the effect of extraversion on novelty seeking and divergent thinking. To test the divergent thinking, the authors used two alternative uses tasks, asking for creative ideas for a "cable" and a "tin can", each to solve within two minutes. They rated the ideas based on originality and further paid every participant a monetary reward of five euros for participating in this study. Their results indicate that novelty seeking is linked to extraversion and can further lead to greater divergent thinking. A possible limitation could be seen in the short time of two minutes, given for each task, as some individuals may come up with a lot of ideas in a short amount of time, while others might need more time to generate novel ideas. Hence, I assume this might have an impact on the results.

Further, Fulmer and Walker (2015) found evidence for increased creativity and productivity for high extraverted people in less structured tasks, but also state that the difference to less extraverted people is rather small. Similar results can also be approved by Sung and Choi (2009). Other studies (Stewart & Bobko, 1996) also tested the interaction effect of reward payments on the relationship between extraversion and sales performance for sales representatives and equally proposed a higher sensitivity for rewards for extraverted individuals and hence, predicted a positive effect resulting in a better performance.

To conclude, various studies indicate that extraverted individuals might be more reward-sensitive and might perform better than other individuals, when it comes to competition. In this case, rewards would positively influence creative performance. Therefore, this study assumes that similar results might be found testing the moderating effect on the reward-originality relationship in a strategic context.

Hypothesis 3b. Extraversion will positively moderate the relationship between rewards and originality.

To give an overview of all the hypotheses presented above, Figure 1 represents this study's conceptual model showing the relationships being analyzed.

3. Methodology

3.1. Data collection

I collected data over a period of one week using personal connections on the one hand, and by making use of these connections to find more participants. This approach allowed me to be aware of the academic and/or working background the participants in this study had, which also ensured the strategic context of this study. Additionally, the link to this experiment was posted on LinkedIn, an online platform for professional business networking and in several social media groups with students studying business administration, management or strategy.

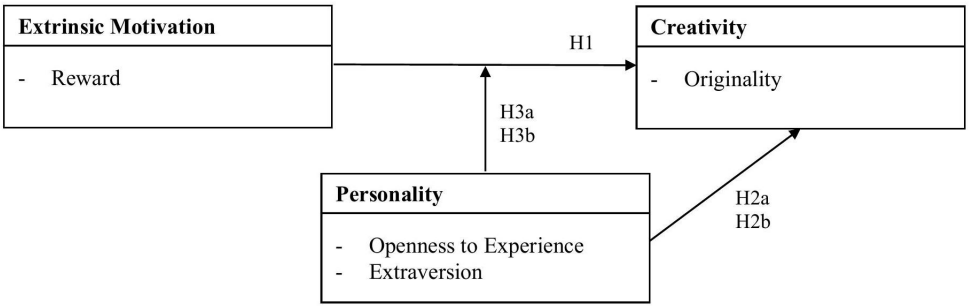


Figure 1: Conceptual Model

Even though this approach allowed me to find a high number of participants, there are still limitations and weaknesses concerning this approach. However, due to this sampling method, I cannot exclude a bias among the participants. First, as this sample was focused on a specific field of participants, it does not represent the population, which should be considered regarding the explanatory power of the results. Second, by using personal connections, I have to assume certain weaknesses in this study regarding participant's motivation to give answers that are more likely to fit into this research, even though the purpose of this study was revealed after finishing the survey and no influencing information was provided beforehand. Third, due to the sharing of the link in social platforms, there was a higher dropout rate than in other sampling methods, which should be taken into account.

3.2. Sample

The sample was composed of 141 participants (no reward condition = 70, reward condition = 71), of which 76 were female, 61 male, and 4 preferred not to say, aged from 18 to 63 years old (M=29.84, SD=9.859). In terms of education, 2 had no high school degree, 9 had some college but no degree, 13 had a high school degree, 55 had a bachelor's degree, 51 had a master's degree, 9 had a PhD or higher, and 2 did not indicate their educational level. Regarding the employment status, 58 were students, 70 were employed, 7 were self-employed and 6 were currently not working. The level of working experience averaged about 7.33 years.

3.3. Setting and Procedure

Pre-test. Several days before distributing the online experiment to participants, I conducted a pre-test to find possible mistakes, unclear formulations and to receive feedback on the survey design. Therefore, three volunteers agreed to do the testing and give feedback. As those volunteers tested both conditions and were informed about the setting, they did not participate in the experiment. Feedback was given especially on the timing for the divergent thinking task, which was supposed to last for 6 minutes but was then, after processing the feedback, reduced to 4 minutes, as the task was described as too long. Additionally, small adaptations were made in the experiment by highlighting some words in the text or adding an explanation for the divergent thinking task on how to type in the answers.

Experiment. The study consisted of two phases, with the personality test first and the divergent thinking test second. Participants completed the study online in Qualtrics, which is an online survey experiment software. I chose to conduct an experiment, as this allows for manipulating variables and then observing the corresponding difference in the outcome - in this study, namely manipulating the rewarding condition, to test possible differences in the creative performance. The experiment lasted approximately 8 minutes.

Participants were randomly assigned to a reward or no-reward expected condition. For the analysis of personality traits this study focused on the Five Factor Model (Costa & McCrae, 1985) using the question repertoire of the Big Five Inventory for openness to experience and extraversion on a 5-point Likert scale (BFI; John and Srivastava, 1999). Before introducing the second part, participants were informed, that the following task was about creativity and that they would be rated for originality by an objective judge. Participants in the reward condition were further informed about the possibility to win a reward. Rewards were of medium size (20€) and were given to the Top 5 participants with the best scores in originality after analyzing the results.

For the second part measuring the divergent thinking ability, this study used an adapted version of the Alternative Uses Task (AUT; Guilford, 1967). Participants were provided with the name of a common object and instructed to think of uncommon or original uses for this object. The time limit to complete the task was four minutes. The task included the following instructions: "*In this task you are asked to use your imagination and be creative. Please note, that there are no wrong answers. Think of all the uses that come into your mind for a fork*". The aim of the task was to find as many original ideas as possible to provide creative solutions.

At the end of the experiment participants were asked to answer some short questions about their enjoyment of the creativity task and some demographic questions. After all the tasks were judged, the 5 best participants were contacted via email on April 16th 2021, to transfer the promised reward.

3.4. Measures

Reward. The independent variable consisted of two different conditions, a no reward condition and a 20 Euro reward condition. The tournament reward was given to the top 5 participants with the best creative answers according to the originality rating. I decided to use such a reward, as the evaluation of each individuals' performance was dependent on the overall performance, which should increase the competitiveness of participants and hence, their creative performance. Tournament rewards are often used to develop innovative ideas and the participation of individuals is often motivated by money, the love for competition or curiosity-seeking (Morgan & Wang, 2010). As this study is focused on creativity, the aim was not to develop innovative outcome, but novel, useful, and creative ideas according to individuals' ability to divergent thinking.

Regarding the size of the reward, several studies indicate that there is no significant effect on performance outcomes, which suggest that a higher reward does not automatically lead to better results (Holst-Hansen & Bergenholtz, 2020; Mason & Watts, 2009). Therefore, this study did not expect any limitations on the validity regarding the size of the reward.

Personality traits. The moderating variable was measured by a short survey. The personality traits asked were Openness to Experience and Extraversion, whereof ten items relate to openness ("has an active imagination", "likes to reflect") and eight items are related to extraversion ("is talkative", "is full of energy"). In total, the personality test consisted of 18 short questions. For the analysis of personality traits this study used an adapted version of the Five Factor Model (Costa & McCrae, 1985), the Big Five Inventory (BFI; John and Srivastava, 1999). According to Chamorro-Premuzic and Reichenbacher (2008, p. 1096) "The BFI is well validated and has internal consistencies ranging from .75 to .90 and test-retest reliabilities ranging from .80 to .90." In the BFI version, respondents state to what extent the elements pertain to them in a scale from 1 (strongly disagree) to 5 (strongly agree). A complete list of the questions can be found in the appendix. The reason why the BFI version was selected is simply because it's shorter than the Costa and McCrae (1985) version.

This study could have also taken other measures into account, to test participants personality. However, empirical research shows that one of the key strengths of the Big Five model is its consistency across time, culture and age as well as its structure, which avoids overlaps (Costa & McCrae, 1992b). Further, it has internal consistency as well as good divergent validity with the traits measured (Myszkowski et al., 2015).

Originality. Divergent thinking tasks are often used in creativity studies and are seen as reliable and valid (Runco & Acar, 2012). The dependent variable in this study was measured by scores for original ideas in the creativity task and ranged from $0-10$ points for each idea. The rating focused on originality, which represents the number of unusual or unique ideas (Runco, 2007). The requirement for each idea to be rated was that it had to be useful, otherwise ideas received 0 points. The ideas were rated by three judges in total.

The procedure of the creativity rating was as follows. First, I created a list with all the ideas participants generated in this task, using MS Excel. For the other two judges it was not possible to link those ideas to participants or any other variables such as age or gender. Second, this list was then distributed to the judges with a brief explanation of the study and clear instructions for the rating. Every judge had to rate the ideas by giving points for originality, as long as the ideas were also useful. Third, after every judge rated every participants' ideas, I added the scores of all ideas per participant together and created the variable originality, based on the sum of those scores. An inter-rater reliability analysis was conducted for the creativity scores of all three judges, indicating to what extent the judges agreed with each other. There is a strong significant correlation $(.0,83)$ between the means of creativity per participant. There were no effects of age or gender on the creativity measure, and also no interaction with the independent variable was given.

Control variables. I also collected data on a number of control variables, such as age, gender, current status of employment, educational level, years of working experience and intrinsic motivation. For measuring intrinsic motivation, the instrument consisted of 3 items using a 7-point scale ranging from 1 (not at all true) to 7 (very true). The questions are based on the Intrinsic Motivation Inventory (IMI; Deci and Ryan, 2003). The questions were (a) "*This activity was fun to do*" (b) "*I would describe this activity as very interesting*", (c) "*I enjoyed doing this activity very much*". These items measured the extent to which participants enjoyed the task and performed it for its own sake.

For analyzing the data and testing the hypothesis IBM SPSS Statistics 27 was used.

4. Analysis and Results

4.1. Missing values and recoding

Before starting the analysis, all data was checked for errors or missing values. As Qualtrics allows a forced answering mode, the questions in this experiment were all completed, and no missing values or data were found. Participants who did not complete the experiment were automatically filtered out by the program and as main parts of the experiment were not completed, those participants - 43 in total, were excluded from the analysis. Additionally, as the Big Five Inventory questionnaire (BFI; John and Srivastava, 1999) for the personality test includes question which ask for reversed scoring, the answers to the question 4, 6, 12, 15 and 18 were reversed before doing the analysis. For further analysis, the variables gender, status of employment, education and working experience were dummy coded.

4.2. Distributions, reliability, and internal consistency

For testing the normal distribution of the variables, I first analyzed the skewness and kurtosis. The skewness is a mea-

sure of the asymmetry of a distribution, whereas the kurtosis measures the extent to which there are outliers. A normal distribution is symmetric and has a skewness value and a kurtosis value of zero. However, the values for both, skewness and kurtosis were analyzed, and some values were computed to normalize the distribution and achieve acceptable values between -1 and +1 (e.g. openness to experience changed from (S=-1.971, SD=.204) to (S=-.976, SD=.204); work experience changed from (S=1.131, SD=.204) to (S=.122, SD=.204). Additionally, to provide more statistical significance on these results, I also conducted a test of normality by doing a Kolmogorov-Smirnov analysis and a Shapiro-Wilk analysis.

To test the internal consistency of the items presented in the personality scale (BFI; John and Srivastava, 1999) and the intrinsic motivation scale (IMI; Deci and Ryan, 2003) I used in my experiment, I tested the effect of each item of those scales using Cronbach's alpha. The calculation of Cronbach's alpha "has become a common practice in (...) research when multiple-item measures of a concept or construct are employed" (Tavakol & Dennick, 2011, p.53). Cronbach's alpha was developed to provide a useful measure of the internal consistency of a test or scales and is presented as a number between 0 and 1, whereas a value above 0,7 is interpreted as consistent. If the items in the measured scale correlate with each other, the alpha value increases. However, a higher value does not automatically stand for a higher degree of internal consistency, as alpha is also influenced by the length of the scale (Tavakol & Dennick, 2011).

In this study, the personality scale had a Cronbach's alpha value of .831 which indicates a high reliability. Additionally, I also tested if deleting items from the personality scale would increase the Cronbach's alpha value. Deleting one item of openness to experience would have led to a higher value of .846 , however I decided to accept the Cronbach's alpha value of .831, as this already indicates a high reliability.

For intrinsic motivation, the scale presented an internal consistency of .826, which is also accepted as a reliable alpha value. However, the results for intrinsic motivation show that the Cronbach's alpha would not increase significantly when deleting a specific item.

I also checked to see whether there were other interesting findings regarding the means, standard deviations or other values for each variable. Regarding the co-variates results revealed that for openness to experience (M=36.16, SD=6.360)[1] and extraversion (M=28.76, SD=5.861) the medians were relatively high, as the maximum scoring for openness to experience was 50 whereas for extraversion it was 40 . An in-depth analysis proposed further that the lowest score for openness to experience was 22 and the highest score was 49, whereas for extraversion the lowest score was 9 and the highest score was 39. However, even though the manipulation of this experiment was after measuring the personality levels, analysis revealed significant results (r=.548, p=.001) indicating that especially for the reward condition individuals showed interestingly higher levels of openness to experience (M=36.99, SD=6.649) and extraversion (M=29.32, SD=5.613) compared to the no reward condition (openness to experience (M=35.33, SD=5.985); extraversion (M=28.19, SD=6.089).

[1] For further clarification it might be useful to note that the personality scores of individual participants were added up and therefore, the sum was taken here instead of the mean. However, statistically this makes no difference.

4.3. Divergent thinking task

The creativity rating revealed that on average, participants had approximately 8 creative ideas on how to use a fork, ranging from 1-30 ideas. In the reward condition participants produced on average 7 ideas, while in the non-reward condition participants produced approximately 9 ideas. The median for the creativity rating in the reward condition was 44.97 (with the lowest possible total score being 5 and the highest possible total score 101), the median for the non-reward condition was 37.00 (with the lowest possible total score being 6 and the highest possible total score 124).

Regarding the rating, in total, 1,214 ideas were generated, whereof 49.01 % received a score of 5 or above. The scores 9 and 10 were not given at all, as the highest given score was 8, received by 0.34 % of the participants, followed by a score of 7(6.84 %) and 6(18.95 %). Ideas which scored over 7 were for instance "use to make patterns in the sand", "use as a zen garden tool", "use to make a sculpture", "use as an ice cream stick", "use as a small ladder for bugs" or "use for gardening".

4.4. Descriptive statistics

Table 1 presents the scale means, standard deviations, intercorrelations, and reliabilities for each of this study's variables. As the table displays, intrinsic motivation is positively correlated with openness to experience (r=.195, p=.020), extraversion (r=.218, p=.009) and originality (r=.213, p=.011). This indicates that those individuals with high levels of openness and extraversion were also highly intrinsically motivated while performing the experiment. Additionally, those individuals intrinsically motivated performed better in the divergent thinking task. For originality, the table shows a significant negative correlation for age ($r = -.187$, p=.027) which means that the older individuals, the less creative they performed. Regarding the working experience, there is also a significant negative correlation to age (r=-.174, p=.039). This implies that there is no relation as work experience is a negative indicator for creativity, as more experience did not lead to more creativity.

4.5. Extrinsic and intrinsic motivation

Further analysis revealed significant results (F=6.366, p=.013) regarding the effect of intrinsic motivation on the reward condition (M=4.75, SD=1.471) and no reward condition (M=4.63, SD=1.625). Results show that the intrinsic motivation for the reward condition was even higher than for the no reward condition and as indicated above, the quality

Table 1: Overall means, standard deviations, and correlations

Variable	*M*	*SD*	1	2	3	4	5	6
1. Openness to Experience	36.16	6.36	-					
2. Extraversion	28.76	5.86	.548*	-				
3. Originality	41.01	19.32	.409**	.451*	-			
4. Intrinsic Motivation	4.69	1.56	.195*	.218**	.213*	-		
5. Age	29.84	9.86	.000	−.181*	−.187*	−.144	-	
6. Work Experience	7.33	9.38	.049	−.160	−.174*	−.148	.921**	-

Note: N=141, **p<.01, *p<.05

of answers regarding its originality was higher in the reward condition. As this assumption is part of an ongoing discussion in the creativity field, this finding is of particular interest and will be discussed in the next section.

4.6. Hypotheses testing

In a first step I analyzed the overall model to test whether it had explanatory power and how much. Results indicated significant results regarding this study's model with an explanatory power of 56.2 % (Rˆ2=.562, F=16.217, p=.001). In a next step, I tested the hypotheses using a regression analysis. Table 2 presents the findings for this study's hypotheses.

Hypothesis 1 proposed that monetary rewards for creativity are positively related to originality. Table 2 shows that the relationship between monetary rewards and originality is statistically not significant (F=.229, p=.633). According to these results, hypothesis 1 is therefore not supported. However, there is still reason to believe that there is a positive relationship between rewards and originality. When testing only those two variables (mean originality score for the no reward condition (M=37.00), mean originality score for the reward condition (M=44.97)), the relationship is significantly positive (F=6.228, p=.014), which indicates that openness to experience and extraversion might have a strong link to originality. This leads to an overlap between those variables which absorbs the explanatory power of this relationship. Additionally, as a further explanation, when conducting the experiment, the personality test was filled in before the manipulation was introduced, which indicates that the results for personality are based on randomness and not biased by the extrinsic reward manipulation. Therefore, this study assumes a disturbing effect between rewards and originality as presented in this model and believes that due to the absorbing power of the two personality characteristics, the relationship between rewards and originality is displayed as not significant, even though it actually is significant in a bivariate analysis.

Hypothesis 2a proposed that openness to experience will be positively related to originality. As displayed in the table above, the results indicate a significantly positive relationship for openness to experience and originality (F=5.300, p=.023). Hypothesis 2a is therefore supported.

Hypothesis 2b proposed that extraversion will be positively related to originality. Table 2 shows that the relationship between extraversion and originality is positive and statistically significant (F=12.856, p=.000). Hypothesis 2b is therefore supported.

Hypothesis 3a proposed that openness to experience will positively moderate the relationship between rewards and originality. Table 2 shows that there is no significant interaction effect of openness to experience with extrinsic rewards (F=.241, p=.624). According to these findings hypothesis 3a is not supported.

Hypothesis 3b proposed that extraversion will positively moderate the relationship between rewards and originality. As Table 2 presents, there is no significant interaction effect of extraversion with extrinsic rewards (F=.073, p=.788). Hypothesis 3b found no support.

5. Discussion

5.1. Discussion of results

The main purpose of this study was to test if personality has a moderating effect on the relationship between extrinsic monetary rewards and creative performance, measured as originality. First, I measured the effect of extrinsic rewards on originality. Second, I tested the effect of openness to experience and extraversion on originality and third, I tested for the moderating effect.

This study proposed a positive effect of extrinsic rewards on originality. Despite the popularity of using extrinsic motivators such as performance-contingent rewards to enhance creativity, there is still little agreement on the effect on creativity (Shalley et al., 2004). Nonetheless, this assumption is also in line with the study of Eisenberger and Rhoades (2001) which tested the effect of rewards on creativity and reported higher results for those participants being rewarded, and the studies of Byron and Khazanchi (2010) or Joussemet and Koestner (1999). Even though the measured means for both conditions implied high significance when testing only for those two variables, the effect was absorbed by the two personality traits when testing the significance of all relevant variables. For hypothesis 1 this indicates that even though the hypothesis is officially not supported, there is still reason to believe that there is a positive relationship. Hence, this

Table 2: Results of the regression analysis for originality

Variable	*Sum of Squares*	*df*	Mean Square	F	P
Reward	65.135	1	65.135	.229	.633
Openness to Experience	1506.159	1	1506.159	5.300	.023
Extraversion	3653.013	1	3653.013	12.856	.000
Moderator_Openness to Experience	68.615	1	68.615	.241	.624
Moderator_Extraversion	20.664	1	20.664	.073	.788
Error	38361.135	135	284.157		
Total	289421.000	141			

shows that an important factor of increasing creativity can be found in extrinsic motivation.

Regarding the impact of personality on divergent thinking, the findings of this study are in line with research in this area proposing positive effects of openness to experience and extraversion on originality. Especially for openness to experience this study was expecting a positive relationship with creativity as this personality trait is known as being "the most strongly tied to creativity" (Runco, 2007, p. 296) and is predicted as being a key personality for creativity (Herrmann & Nadkarni, 2014; Nadkarni & Herrmann, 2010; H. Zhao & Seibert, 2006; H. Zhao et al., 2010). Research indicates that openness to experience is necessary to explore creative solutions to problems which is also in line with the study findings of Myszkowski et al. (2015). Additionally, no research was found stating the opposite, which indicates that openness to experience and creativity are strongly linked to each other. Therefore, there was no reason to expect a different outcome in a management context, which is approved by these results.

For extraversion, research proposes that individuals with a high level of extraversion tend to be enthusiastic and ambitious, which results in a proactive behavior by actively engaging in tasks and trying to find novel ideas (Raja et al., 2004). This is also supported by Gocłowska et al. (2019) stating that extraversion is not only positively related to creativity, but also to novelty seeking which leads to greater divergent thinking. In addition, Chamorro-Premuzic and Reichenbacher (2008) argue that extraverted individuals have an intrinsic motivated advantage in divergent thinking tasks especially when they are being evaluated. This is also displayed in this study's findings presenting a positive significant correlation between extraversion and intrinsic motivation. Therefore, extraversion can be assumed as facilitator for divergent thinking and predicts creativity.

Finally, this study tested the moderating effect of personality on the relationship between extrinsic rewards and originality. Existing research indicates high reward-sensitivity for individuals with high levels of openness to experience (Vandenberghe et al., 2008) and extraversion (Ashton et al., 2002; Lucas & Diener, 2001). Unfortunately, this study found no support for a moderating effect for both personality characteristics. A possible explanation could be that even though research indicates that higher rewards do not automatically lead to better results (e.g. Mason and Watts, 2009), this study assumes that higher rewards might have led to more significantly positive results regarding the moderation effect. A different and more reasonable explanation could be that personality characteristics simply do not act as moderators in this relationship which sets the stage for future research in this area.

5.2. Theoretical and managerial Implications

This study contributes to existing literature in several ways. First, openness to experience and extraversion were both shown to have a positive effect on creativity, which is consistent with the literature. This is therefore not a new finding as such, but - to the extent that this study allows - it reinforces the validity of these assumptions.

Second, regarding the intrinsic and extrinsic motivation, there is an ongoing debate in the literature which has resulted in contradictory findings. According to Amabile (1998) people are most creative, when they feel intrinsically motivated, while Joussemet and Koestner (1999) claim that extrinsic rewards undermine intrinsic motivation. In fact, the results of this study revealed that there was a higher intrinsic motivation for the reward condition than for the no reward condition. As this study found even higher intrinsic motivation in the rewarding condition, I could not find an undermining effect resulting in less creative performance. This is also approved by Bradler et al. (2016) claiming that tournaments indeed increase creative performance, without any evidence for crowding out intrinsic motivation. In my opinion, this finding resulted due to the high competitiveness among participants in order to perform better or even outperform others, which may have also increased their intrinsic motivation. However, despite extensive research I could not find any literature-based explanation for this finding, which provides a very interesting research idea for future investigations. Thus, this finding indicates that the intrinsic and extrinsic motivation of individuals can be balanced when performing a creative task, without one motivation undermining the other and therefore contributes to the ongoing discussion.

Additionally, when focusing on the extrinsic motivation, this study also indicated that creative performance was higher when rewarding creativity. In practical terms, this finding can be used by managers for instance to conduct idea tournaments to unlock a companies' innovative poten-

tial by motivating employees to participate and reward them for their creative ideas (Morgan & Wang, 2010).

Lastly, the main focus of this study was to test the moderating effect of personality on the reward-creativity relationship. As companies need creative, and motivated employees in their teams, the question arises how this can be achieved. Therefore, this study wanted to find out if openness to experience and extraversion affect this relationship, which could have served as guidance for managers in terms of employee appraisals, recruiting interviews, or promotions. However, even though the moderating effect was not supported, the findings offer useful insights for practice. The study confirmed that openness to experience and extraversion, both have an impact on creative performance. This is interesting for companies in regards to team or project formation, to find the right balance. As some companies already test their employees' personality before joining a project team (e.g. Deloitte), this finding can serve as additional guidance to create and maintain diversity, as this is essential for creativity (Kurtzberg & Amabile, 2001).

5.3. Limitations and future research

This study has several limitations which provide a stage for future research. First, this study was conducted as an online experiment for individuals with a (strategic) management background. Even though in total I had 141 participants, I assume that the validity is a different one compared to experiments conducted in other settings, such as field experiments with less participants. The results concerning the effect of rewards on creativity or the moderating effect of personality may have been different in this setting. Due to the current pandemic situation there was no other choice of conducting the experiment, however, future research could focus its' analysis conducting a field experiment to provide more realistic results.

Second, the results obtained for the personality characteristics and for the intrinsic motivation were self-reported. This leads to the assumption that many participants may have a different self-image and therefore answered accordingly. This study thus assumes that participants' responses may be biased due to self-assessment. Future research could have the personality survey conducted by objective and independent third parties to avoid these same biases, which would lead to better results.

Third, an important variable in this experiment was the introduction of the manipulation in the form of an extrinsic reward. The aim was not only to determine the effect of extrinsic motivation on creativity, but also to find out whether the two personality factors openness to experience and extraversion react more sensitively to extrinsic rewards, and thus whether a statement can be made about the interaction effect. This hypothesis could not be confirmed, which suggests that the limitation lies in the amount of reward. However, although I have referred to studies suggesting the opposite (e.g. Holst-Hansen and Bergenholtz, 2020; Mason and Watts, 2009), it is reasonable to assume that the possibility of a higher reward would have allowed a clearer statement regarding the interaction with personality. Future research should consider the possibility of introducing rewards at different levels (e.g. 1 € , 10 € , 25 €), and test each participant through each condition, in order to have clearer results in combination with the personality characteristics.

5.4. Conclusion

The aim of this study was to examine the effect of openness to experience and extraversion on the relationship between extrinsic rewards and originality. First, even though in the findings the effect of rewards on originality was not significant, there is still reason to believe that extrinsic rewards positively influence creative performance. Second, it was confirmed that both personality characteristics are positively related to creativity, which indicates that individuals with high levels of openness to experience or extraversion also achieve higher creative performance. Unfortunately, this study could not find support for a moderating effect of those personality traits. However, the results of this study also revealed that there was a higher intrinsic motivation of individuals for the reward condition than for the no reward condition, which indicates that the intrinsic and extrinsic motivation of individuals can be balanced when performing a creative task, without one motivation undermining the other. Additionally, it shows the impact of tournament rewards on motivation and creativity, which sets a stage for future research. Nevertheless, the results of this study contribute to the ongoing discussion in the creativity field regarding extrinsic motivators and provides useful insights for managers concerning personality traits and how to use them effectively to increase creativity.

References

Amabile, T. M. (1983). The social psychology of creativity: A componential conceptualization. *Journal of Personality and Social Psychology*, *45*(2), 357–376. https://doi.org/10.1037/0022-3514.45.2.357

Amabile, T. M. (1993). Motivational synergy: Toward new conceptualizations of intrinsic and extrinsic motivation in the workplace. *Human Resource Management Review*, *3*(3), 185–201. https://doi.org/10.1016/1053-4822(93)90012-S

Amabile, T. M. (1996). *Creativity in context*. Boulder, CO: Westview.

Amabile, T. M. (1998). How to kill creativity. *Harvard Business Review*, *76*(5), 77–87.

Amabile, T. M., & Gryskiewicz, N. D. (1989). The creative environment scales: Work environment inventory. *Creativity Research Journal*, *2*(4), 231–252. https://doi.org/10.1080/10400418909534321

Amabile, T. M., Hadley, C. N., & Kramer, S. J. (2002). Creativity under the gun. *Harvard Business Review*, *80*(8), 52–61.

Ames, M., & Runco, M. (2005). Predicting Entrepreneurship From Ideation and Divergent Thinking. *Creativity and Innovation Management*, *14*(3), 311–315. https://doi.org/10.1111/j.1467-8691.2004.00349.x

Ashton, M., Lee, K., & Paunonen, S. (2002). What is the central feature of extraversion? Social attention versus reward sensitivity. *Journal of Personality and Social Psychology*, *83*(1), 245–251. https://doi.org/10.1037/0022-3514.83.1.245

Baer, M., Oldham, G. R., & Cummings, A. (2003). Rewarding creativity: When does it really matter? *Leadership Quarterly*, *14*(4), 569–586. https://doi.org/10.1016/S1048-9843(03)00052-3

Barrick, M. R., & Mount, M. K. (1991). The Big Five Personality Dimensions and Job Performance: A Meta-Analysis. *Personnel Psychology*, *44*(1), 1–26. https://doi.org/10.1111/j.1744-6570.1991.tb00688.x

Barron, F., & Harrington, D. (1981). Creativity, Intelligence, and Personality. *Annual Review of Psychology*, *32*(1), 439–476. https://doi.org/10.1146/annurev.ps.32.020181.002255

Behrens, J., Ernst, H., & Shepherd, D. A. (2014). The decision to exploit an R&D project: Divergent Thinking across Middle and Senior Managers. *The Journal of Product Innovation Management*, *31*(1), 144–158. https://doi.org/10.1111/jpim.12085

Biraglia, A., & Kadile, V. (2017). The role of entrepreneurial passion and creativity in developing entrepreneurial intentions: insights from american homebrewers. *Journal of Small Business Management*, *55*, 170–188. https://doi.org/10.1111/jsbm.12242

Blain, S. D., Sassenberg, T. A., Xi, M., Zhao, D., & DeYoung, C. G. (2021). Extraversion but not depression predicts reward sensitivity: Revisiting the measurement of anhedonic phenotypes. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/pspp0000371

Bonner, S. E., Hastie, R., Sprinkle, G. B., & M., Y. S. (2000). A review of the effects of financial incentives on performance in laboratory tasks: Implications for management accounting. *Journal of Management Accounting Research*, *12*(1), 19–64. https://doi.org/10.2308/jmar.2000.12.1.19

Bradler, C., Neckermann, S., & Warnke, A. (2016). Incentivizing Creativity: A Large-Scale Experiment with Tournaments and Gifts. *Tinbergen Institute Discussion Paper*.

Byron, K., & Khazanchi, S. (2010). When and How rewards increase creative performance: A theoretically-derived meta-analysis. *Annual Meeting Proceedings (Academy of Management)*, *8*(1), 1–6. https://doi.org/10.5465/AMBPP.2010.54485861

Byron, K., & Khazanchi, S. (2012). Rewards and creative performance: a meta-analytic test of theoretically derived hypotheses. *Psychological Bulletin*, *138*(4), 809–830. https://doi.org/10.1037/a0027652

Chamorro-Premuzic, T., & Reichenbacher, L. (2008). Effects of Personality and Threat of Evaluation on Divergent and Convergent Thinking. *Journal of Research in Personality*, *42*(4), 1095–1101. https://doi.org/10.1016/j.jrp.2007.12.007

Costa, P. T., & McCrae, R. R. (1985). *The NEO Personality Inventory manual*. Odessa, FL: Psychological Assessment Resources.

Costa, P. T., & McCrae, R. R. (1992a). Four ways five factors are basic. *Personality and Individual Differences*, *13*(6), 653–665. https://doi.org/10.1016/0191-8869(92)90236-I

Costa, P. T., & McCrae, R. R. (1992b). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.

Dai, S., Li, Y., & Zhang, W. (2019). Personality traits of entrepreneurial top management team members and new venture performance. *Social Behavior and Personality: An international journal*, *47*(7), 1–17. https://doi.org/10.2224/sbp.8107

Deci, E. L., Koestner, R., & Ryan, R. M. (2001). Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. *Review of Educational Research*, *71*, 1–27. https://doi.org/http://doi.org.proxy.uba.uva.nl:2048/cv9qch

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. https://doi.org/10.1016/0092-6566(85)90023-6

Deci, E. L., & Ryan, R. M. (2003). Intrinsic Motivation Inventory. http://selfdeterminationtheory.org/edu/scales/category/5-intrinsic-motivation-inventory

Depue, R. A., & Collins, P. F. (1999). Neurobiology of the structure of personality: Dopamine facilitation of incentive motivation and extraversion. *Behavioral and Brain Sciences*, *22*, 491–569. https://doi.org/10.1017/S0140525X99002046

Dollinger, S., Urban, K. K., & James, T. A. (2004). Creativity and Openness: Further Validation of Two Creative Product Measures. *Creativity Research Journal*, *16*(1), 35–47. https://doi.org/10.1207/s15326934crj1601_4

Eisenberger, R. (1992). Learned industriousness. *Psychological Review*, *99*(2), 248–267. https://doi.org/10.1037/0033-295X.99.2.248

Eisenberger, R., & Rhoades, L. (2001). Incremental Effects of Reward on Creativity. *Journal of Personality and Social Psychology*, *81*(4), 728–741. https://doi.org/10.1037/0022-3514.81.4.728

Fulmer, I. S., & Walker, W. J. (2015). More Bang for the Buck? : Personality Traits as Moderators of Responsiveness to Pay-for-Performance. *Human Performance*, *28*(1), 40–65. https://doi.org/10.1080/08959285.2014.974755

George, J. M., & Zhou, J. (2001). When openness to experience and conscientiousness are related to creative behavior: An interactional approach. *Journal of Applied Psychology*, *86*, 513–524. https://doi.org/10.1037/0021-9010.86.3.513

George, J. M., & Zhou, J. (2002). Understanding when bad moods foster creativity and good ones don't: The role of context and clarity of feelings. *Journal of Applied Psychology*, *87*, 687–697. https://doi.org/10.1037/0021-9010.87.4.687

Gocłowska, M. A., Ritter, S. M., Elliot, A. J., & Baas, M. (2019). Novelty seeking is linked to openness and extraversion and can lead to greater creative performance. *Journal of Personality*, *87*(2), 252–266. https://doi.org/10.1111/jopy.12387

Guilford, J. P. (1956). The structure of intellect. *Psychological Bulletin*, *53*(4), 267–293.

Guilford, J. P. (1967). *The nature of human intelligence*. New York: McGraw Hill.

Gupta, N., & Shaw, J. D. (2014). Employee compensation: the neglected area of HRM research. *Human Resource Management Review*, *24*, 1–4. https://doi.org/10.1016/j.hrmr.2013.08.007

Hackman, J. R., & Oldham, G. R. (1980). *Work redesign*. Reading, MA: Addison-Wesley.

Harrison, J. S., Thurgood, G. R., Boivie, S., & Pfarrer, M. D. (2019). Measuring CEO personality: Developing, validating, and testing a linguistic tool. *Strategic Management Journal*, *40*(8), 1316–1330. https://doi.org/10.1002/smj.3023

Heineck, G. (2011). Does it pay to be nice? Personality and earnings in the United Kingdom. *Industrial & Labor Relations Review*, *64*(5), 1020–1038. https://doi.org/10.1177/001979391106400509

Herrmann, P., & Nadkarni, S. (2014). Managing Strategic Change: The Duality of CEO Personality. *Strategic Management Journal*, *35*(9), 1318–1342. https://doi.org/10.1002/smj.2156

Holst-Hansen, J., & Bergenholtz, C. (2020). Does the size of rewards influence performance in cognitively demanding tasks? *PloS One*, *15*(10), e0240291. https://doi.org/10.1371/journal.pone.0240291

Huo, K. (2020). Performance Incentives, Divergent Thinking Training, and Creative Problem Solving. *Journal of Management Accounting Research*, *32*(1), 159–176. https://doi.org/10.2308/jmar-52479

John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In Pervin, L. A. and John, O. P. (Ed.), *Handbook of personality: Theory and research* (2nd ed., pp. 102–138). New York: Guilford Press.

Joussemet, M., & Koestner, R. (1999). Effect of Expected Rewards on Children's Creativity. *Creativity Research Journal*, *12*(4), 231–239. https://doi.org/10.1207/s15326934crj1204_1

Judge, T. A., & Bono, J. E. (2000). Five-factor model of personality and transformational leadership. *Journal of Applied Psychology*, *85*(5), 751–765. https://doi.org/10.1037/0021-9010.85.5.751

Kachelmeier, S. J., Reichert, B. E., & Williamson, M. G. (2008). Measuring and motivating quantity, creativity, or both. *Journal of Accounting Research*, *46*(2), 341–373. https://doi.org/10.1111/j.1475-679X.2008.00277.x

Kaspi-Baruch, O. (2019). Big Five Personality and Creativity: The Moderating Effect of Motivational Goal Orientation. *The Journal of Creative Behavior*, *53*(3), 325–338. https://doi.org/10.1002/jocb.183

Kaufman, S. B., Quilty, L. C., Grazioplene, R. G., & Hirsh, J. B. (2015). Openness to experience and intellect differentially predict creative achievement in the arts and sciences. *Journal of Personality*, *84*, 248–258. https://doi.org/10.1111/jopy.12156

King, L. A., Walker, L. M., & Broyles, S. J. (1996). Creativity and the Five-Factor Model. *Journal of Research in Personality*, *30*, 198–203. https://doi.org/10.1006/jrpe.1996.0013

Kohn, A. (1993). Why Incentive Plans Cannot Work. *Harvard Business Review*, *71*(5).

Kruglanski, A. W., Friedman, I., & Zeevi, G. (1971). The effects of extrinsic incentive on some qualitative aspects of task performance. *Journal of Personality*, *39*, 606–617.

Kurtzberg, T. R., & Amabile, T. M. (2001). From Guilford to creative synergy: Opening the black box of team-level creativity. *Creativity Research Journal*, *13*(3-4), 285–294.

Li, F., Deng, H., Leung, K., & Zhao, Y. (2017). Is Perceived Creativity-Reward Contingency Good for Creativity? The Role of Challenge and Threat Appraisals. *Human Resource Management*, *56*(4), 693–709. https://doi.org/10.1002/hrm.21795

Li, W., Li, X., Huang, L., Kong, X., Yang, W., Wei, D., & Li, J. (2015). Brain structure links trait creativity to openness to experience. *Social Cognitive and Affective Neuroscience*, *10*(2), 191–198. https://doi.org/10.1093/scan/nsu041

Lubart, T. (2016). Creativity and Convergent Thinking: Reflections, Connections and Practical Considerations. *Journal of Psychology and Pedagogics*, *4*, 7–15. https://doi.org/10.22363/2313-1683-2016-4-7-15

Lucas, R. E., & Diener, E. (2001). Understanding Extraverts' Enjoyment of Social Situations: The Importance of Pleasantness. *Journal of Personality and Social Psychology*, *81*(2), 343–356. https://doi.org/10.1037/0022-3514.81.2.343

Madjar, N., Oldham, G. R., & Pratt, M. G. (2002). There's no place like home? The contributions of work and nonwork creativity support to employees' creative performance. *Academy of Management Journal*, *45*, 757–767. https://doi.org/10.2307/3069309

Malik, M. A. R., Butt, A. N., & Choi, J. N. (2015). Rewards and employee creative performance: Moderating effects of creative self-efficacy, reward importance, and locus of control. *Journal of Organizational Behavior*, *36*(1), 59–74. https://doi.org/10.1002/job.1943

Malik, M. A. R., Choi, J. N., & Butt, A. N. (2019). Distinct effects of intrinsic motivation and extrinsic rewards on radical and incremental creativity: The moderating role of goal orientations. *Journal of Organizational Behavior*, *40*(9-10), 1013–1026. https://doi.org/10.1002/job.2403

Mason, W., & Watts, D. (2009). Financial incentives and the "performance of crowds". *Proceedings of the ACM SIGKDD Workshop on Human Computation*, 77–85. https://doi.org/10.1145/1600150.1600175

McAuliffe, M. (2016). The Potential Benefits of Divergent Thinking and Metacognitive Skills in STEAM Learning: A discussion paper. *International Journal of Innovation, Creativity and Change*, *2*(3), 71–82.

McCrae, R. R. (1987). Creativity, divergent thinking, and openness to experience. *Journal of Personality and Social Psychology*, *52*, 1258–1265. https://doi.org/10.1037/0022-3514.52.6.1258

Morgan, J., & Wang, R. (2010). Tournaments for ideas. *California Management Review*, *52*(2), 77–97.

Mumford, M. (2000). Managing Creative People: Strategies and Tactics for Innovation. *Human Resource Management Review*, *10*(3), 313–351. https://doi.org/10.1016/S1053-4822(99)00043-1

Myszkowski, N., Storme, M., Davila, A., & Lubart, T. (2015). Managerial creative problem solving and the Big Five personality traits. *The Journal of Management Development*, *34*(6), 674–684. https://doi.org/10.1108/JMD-12-2013-0160

Nadkarni, S., & Herrmann, P. (2010). CEO Personality, Strategic Flexibility, and firm performance: The case of the Indian Business Process Outsourcing Industry. *Academy of Management Journal*, *53*(5), 1050–1073. https://doi.org/10.5465/amj.2010.54533196

Oldham, G. R., & Cummings, A. (1996). Employee creativity: Personal and contextual factors at work. *Academy of Management Journal*, *39*, 607–634. https://doi.org/10.2307/256657

Perry-Smith, J., & Mannucci, P. (2017). From Creativity to Innovation: The Social Network Drivers of the Four Phases of the Idea Journey. *The Academy of Management Review*, *42*(1), 53–79. https://doi.org/10.5465/amr.2014.0462

Peterson, R. S., Smith, D. B., Martorana, P. V., & Owens, P. D. (2003). The Impact of Chief Executive Officer Personality on Top Management Team Dynamics: One Mechanism by Which Leadership Affects Organizational Performance. *Journal of Applied Psychology*, *88*(5), 795–808. https://doi.org/10.1037/0021-9010.88.5.795

Puryear, J. S., Kettler, T., & Rinn, A. N. (2017). Relating Personality and Creativity: Considering What and How We Measure. *The Journal of Creative Behavior*, *53*(2), 232–245. https://doi.org/10.1002/jocb.174

Raja, U., Johns, G., & Ntalianis, F. (2004). The Impact of Personality on Psychological Contracts. *Academy of Management Journal*, *47*(3), 350–367. https://doi.org/10.2307/20159586

Runco, M. (2007). *Creativity: theories and themes: research, development, and practice*. Elsevier Academic Press.

Runco, M., & Acar, S. (2012). Divergent Thinking as an Indicator of Creative Potential. *Creativity Research Journal*, *24*(1), 1–10. https://doi.org/10.1080/10400419.2012.652929

Runco, M., Kaufmann, J. C., Smith, J. K., & Smith, L. F. (2008). Commentary: Divergent Thinking Is Not Synonymous With Creativity. *Psychology of Aesthetics, Creativity, and the Arts*, *2*(2), 93–96. https://doi.org/10.1037/1931-3896.2.2.93

Saether, E. (2020). Creativity-Contingent Rewards, Intrinsic Motivation, and Creativity: The Importance of Fair Reward Evaluation Procedures. *Frontiers in Psychology*, *11*, 974. https://doi.org/10.3389/fpsyg.2020.00974

Scratchley, L. S., & Hakstian, A. R. (2001). The measurement and prediction of managerial creativity. *Creativity Research Journal*, *13*, 367–384. https://doi.org/10.1207/S15326934CRJ1334_14

Shalley, C., Zhou, J., & Oldham, G. R. (2004). The Effects of Personal and Contextual Characteristics on Creativity: Where Should We Go from Here? *Journal of Management*, *30*(6), 933–958. https://doi.org/10.1016/j.jm.2004.06.007

Shane, S., Nicolaou, N., Cherkas, L., & Spector, T. D. (2010). Do openness to experience and recognizing opportunities have the same genetic source? *Human Resource Management*, *49*(2), 291–303. https://doi.org/10.1002/hrm.20343

Shaw, J. D., Duffy, M. K., Mitra, A., Lockhart, D. E., & Bowler, M. (2003). Reactions to merit pay increases: A longitudinal test of a signal sensitivity perspective. *Journal of Applied Psychology*, *88*, 538–544. https://doi.org/10.1037/0021-9010.88.3.538

Shi, B., David, Y. D., & Yongli, L. (2016). Openness to Experience as a Moderator of the Relationship between Intelligence and Creative Thinking: A Study of Chinese Children in Urban and Rural Areas. *Frontiers in Psychology*, *7*, 641. https://doi.org/10.3389/fpsyg.2016.00641

Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T., Hess, K. I., Martinez, J. L., & Richard, C. A. (2008). Assessing creativity with divergent thinking tasks: Exploring the reliability and validity of new subjective scoring methods. *Psychology of Aesthetics, Creativity, and the Arts*, *2*(2), 68–85. https://doi.org/10.1037/1931-3896.2.2.68

Smillie, L. D., Cooper, A. J., Wilt, J., Revelle, W., & King, L. (2012). Do Extraverts Get More Bang for the Buck? Refining the Affective-Reactivity Hypothesis of Extraversion. *Journal of Personality and Social Psychology*, *103*(2), 306–326. https://doi.org/10.1037/a0028372

Speed, B., Nelson, B., Levinson, A., Perlman, G., Klein, D., Kotov, R., & Hajcak, G. (2018). Extraversion, neuroticism, and the electrocortical response to monetary rewards in adolescent girls. *Biological Psychology*, *136*, 111–118. https://doi.org/10.1016/j.biopsycho.2018.05.017

Stewart, G. L., & Bobko, P. (1996). Reward structure as a moderator of the relationship between extraversion and sales performance. *Journal of Applied Psychology*, *81*, 619–627. https://doi.org/10.1037/0021-9010.81.6.619

Sue-Chan, C., & Hempel, P. S. (2016). The Creativity-Performance Relationship: How Rewarding Creativity Moderates the Expression of Creativity. *Human Resource Management*, *55*(4), 637–653. https://doi.org/10.1002/hrm.21682

Sung, S. Y., & Choi, J. N. (2009). Do Big Five Personality Factors affect individual creativity? The moderating role of extrinsic motivation.

Social Behavior and Personality: An International Journal, *37*(7), 941–956. https://doi.org/10.224/sbp.2009.37.7.941

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Tierney, P., & Farmer, S. M. (2002). Creative self-efficacy: Its potential antecedents and relationship to creative performance. *Academy of Management Journal*, *45*, 1137–1148. https://doi.org/10.2307/3069429

Vandenberghe, C., St-Onge, S., & Robineau, É. (2008). An Analysis of the Relation between Personality and the Attractiveness of Total Rewards Components. *Relations Industrielles*, *63*(3), 425–453. https://doi.org/10.7202/019096ar

Walker, B. R., & Jackson, C. J. (2004). How the Five Factor Model and revised Reinforcement Sensitivity Theory predict divergent thinking. *Personality and Individual Differences*, *57*, 54–58. https://doi.org/10.1016/j.paid.2013.09.011

Webb, R. A., Williamson, M. G., & Zhang, Y. M. (2013). Productivity-target difficulty, target-based pay, and outside-the-box thinking. *The Accounting Review*, *88*(4), 1433–1457. https://doi.org/10.2308/accr-50436

Weiss, S., Steger, D., Kaur, Y., Hildebrandt, A., Schroeders, U., & Wilhelm, O. (2020). On the Trail of Creativity: Dimensionality of Divergent Thinking and its Relation with Cognitive Abilities, Personality, and Insight. *European Journal of Personality*, 1–22. https://doi.org/10.1002/per.2288

Wu, A. D., & Zumbo, B. D. (2008). Understanding and Using Mediators and Moderators. *Social indicators research*, *87*(3), 367–392. https://doi.org/10.1007/s11205-007-9143-1

Yoon, H. J., Sung, S. Y., & Choi, J. N. (2015). Mechanisms Underlying Creative Performance: Employee Perceptions of Intrinsic and Extrinsic Rewards for Creativity. *Social Behavior and Personality*, *43*(7), 1161–1179. https://doi.org/10.2224/sbp.2015.43.7.1161

Yoon, H. J., Sung, S. Y., Choi, J. N., Lee, K., & Kim, S. (2015). Tangible and Intangible Rewards and Employee Creativity: The Mediating Role of Situational Extrinsic Motivation. *Creativity Research Journal*, *27*(4), 383–393. https://doi.org/10.1080/10400419.2015.1088283

Zhao, H., & Seibert, S. E. (2006). The Big Five personality dimensions and entrepreneurial status: A metanalytical review. *Journal of Applied Psychology*, *91*, 259–271. https://doi.org/10.1037/0021-9010.91.2.259

Zhao, H., Seibert, S. E., & Lumpkin, G. T. (2010). The Relationship of Personality to Entrepreneurial Intentions and Performance: A Meta-Analytic Review. *Journal of Management*, *36*, 381–404. https://doi.org/10.1177/0149206309335187

Zhao, J., Sun, W., Zhang, S., & Zhu, X. (2020). How CEO Ethical Leadership Influences Top Management Team Creativity: Evidence From China. *Frontiers in Psychology*, *11*, 748. https://doi.org/10.3389/fpsyg.2020.00748

Zhou, J. (1998). Feedback valence, feedback style, task autonomy, and achievement orientation: Interactive effects on creative performance. *Journal of Applied Psychology*, *83*(2), 261–276. https://doi.org/10.1037/0021-9010.83.2.261

Zhou, J., & George, J. M. (2001). When job dissatisfaction leads to creativity: Encouraging the expression of voice. *Academy of Management Journal*, *44*, 682–696. https://doi.org/10.2307/3069410

Junior Management Science 9(1) (2024) 1286-1305

Junior Management Science

www.jums.academy
ISSN: 2942-1861

The Effect of Carbon Taxes on Directed Technological Innovation: A Case Study of Sweden

Paolo Oppelt

Technical University of Munich

Abstract

A carbon tax is widely seen as an effective climate policy instrument for discouraging the emission of greenhouse gases that cause climate change. According to the economic theory of the Porter hypothesis, a carbon tax can induce directed technological change toward innovation in clean technologies. Nevertheless, empirical research on the effects of a carbon tax on clean innovation, especially concerning recent periods, is sparse. This paper uses a quasi-experimental approach, in the form of the synthetic control method, to estimate the effect of carbon taxes on climate change mitigating technologies. I conduct a case study of the introduction of the carbon tax in 1990/1991 in Sweden and its effect on clean technology in the transportation sector. Sweden is chosen as it was the first country, next to Finland, to implement a carbon tax, and that at a significant price. I find that the introduction of the carbon tax in 1990/1991 has a positive effect with an economically meaningful magnitude on driving innovation in climate change mitigating technologies. The significant and strong effect of the carbon tax on clean innovation can provide important policy insights for other governments, which did not yet introduce a carbon tax or did not do so at an insignificant rate.

Keywords: carbon taxation; clean technology innovation; Sweden; synthetic control method; transportation sector

1. An Introduction to the Role of Carbon Taxes in Carbon Pricing

Since the middle of the previous century, human activity has caused large amounts of greenhouse gas (GHG) emissions to be emitted into the atmosphere, which caused climate change (Azam et al., 2021; Khan et al., 2021). This anthropogenic climate change is one of the imperative issues of our time (United Nations, 2021). Carbon pricing, in the form of a carbon tax or an emissions trading system (ETS), is regarded by many as the policy of choice to achieve the goal of limiting global warming to below 2°C, preferably even to 1.5°C, which was set forth in the Paris Agreement in 2015 (Commission, 2017).

Some scientists advocate using an ETS over a carbon tax because of its better dynamical performance as a cap-and-trade system. The mechanism underlying an ETS is the following: An ETS puts a cap on the total amount of GHG emissions emitted annually. The entities covered by the system receive, buy, or trade allowances to emit GHG emissions with other market participants. This is done to the degree that the regulated entities obtain enough allowances to cover their emissions. Hence, the price of the ETS allowances is determined market-based through supply and demand. The cap ensures that the required emissions reductions occur and emitters stay within their carbon budgets (European Commission, 2022a; World Bank, 2022).

Other scientists advocate a carbon tax over an ETS-based approach (Metcalf & Weisbach, 2009; Weitzman, 1974). Under a carbon tax, the price per ton of GHG emissions is fixed and set by the government implementing it. Proponents of

I would like to express my sincere gratitude to Christoph Ratay for the significant role he played throughout my master's thesis project. His willingness to dedicate time to listen to my ideas, and the thoughtful suggestions he offered, have been crucial in the development of my work. Christoph's belief in my vision fostered a sense of confidence that was essential to my pursuit of academic excellence. His guidance served as a constant source of inspiration, and for that, I am profoundly thankful.

DOI: https://doi.org/10.5282/jums/v9i1pp1286-1305

a carbon tax argue that a carbon tax allows for fast and precise price control to ensure that carbon is priced at a sufficiently high level. This allows the carbon price to interact more harmoniously with other government policies (Goulder & Schein, 2013).

There is empirical evidence that the ETS drives eco-innovation (Dechezleprêtre & Sato, 2017). Moreover, there exists comparative literature on the advantages and disadvantages of both carbon pricing options (Chen et al., 2020). However, there is less research on whether the carbon tax is a policy instrument that should be used in those sectors that the EU ETS does not cover. More specifically, there is a lack of research regarding the effectiveness of implementing a carbon tax to increase eco-innovation and thus achieve long-term GHG emission reduction. I use the implementation of a carbon tax in Sweden as a case study and subsequently explain to which degree my findings may serve as a recommendation for other countries.

On a global scale, ETSs can predominantly be found in the European Union (EU), China, and parts of the US. The EU ETS is currently the second largest ETS in the world, next to the Chinese ETS, which was launched in 2021 (IISD, 2021). It currently encompasses 30 countries and about 40% of the EU's GHG emissions (European Commission, 2022b). The EU ETS has successfully reduced the emissions of energy-intensive industries, such as the manufacturing, power, and aviation sectors (Borghesi & Montini, 2016; Dechezleprêtre & Sato, 2017; Venmans, 2012). According to the Effort Sharing Regulation (ESR), each EU member state has binding annual GHG emission targets for 2021-2030 regarding those sectors that the EU ETS does not cover (European Commission, 2022a). These sectors are transportation, buildings, agriculture, small industry (non-ETS), and waste. They account for around 60% of GHG emissions. ETS member states principally use national ETSs or a carbon tax to reduce emissions in the sectors not covered by the EU ETS (Life Unify, 2022).

The preeminent argument for using carbon pricing as a mitigation instrument is that polluters are economically incentivized to reduce emissions because the previously unpriced negative externalities, GHG emissions, are now priced. This price signal then triggers the development, innovation, and deployment of technologies that emit lower levels or even zero GHGs. These technologies are coined as clean technologies. The underlying dynamics are discussed more extensively in Chapter 2.1. Another argument for carbon pricing is that the technological transition to ecotechnologies can be accelerated through state intervention. Research on technological transition shows that there are other factors next to the pricing of negative externalities that influence the pace of innovation. Some countries may face carbon lock-in, meaning they prefer to use and develop technologies that emit high levels of carbon, so-called dirty technologies. This is the case because there are both decreasing marginal costs in improving already existing technologies and increasing returns through network effects. Therefore, innovation in new clean technologies is competitively disadvantageous in terms of marginal costs and returns (Unruh, 2000). Intervention through policies like a carbon tax can break up the lock-in and subsequently increase the pace of innovation (Arthur, 1989).

Carbon pricing can trigger short-term and longer-term effects. Short-term effects are characterized by operational changes to existing assets, like switching coal to gas as an input source for power-producing plants (Vogt-Schilb et al., 2018). These short-term carbon reduction techniques reduce emissions immediately but do not foster the necessary transformational change toward long-term or full decarbonization. Longer-term effects of carbon pricing on emission reduction are achieved through innovation in clean technology. Many researchers, for instance, study the long-term effects of carbon pricing by analyzing the directed technological innovation in clean technology rather than the shorter-term emission reduction level (Vogt-Schilb et al., 2018). Hence, in this paper, I investigate the effect of carbon pricing, in terms of a carbon tax, on directed technological change toward clean technology innovation. *My research question is: What is the impact of introducing a carbon tax on clean innovation?*

In order to answer my research question, I choose to conduct a case study of Sweden and the transportation sector and then discuss to which degree these findings can be applied to other countries and sectors. The main reason Sweden is chosen is that Sweden was one of the first countries globally to introduce a carbon tax and that at a high rate. The transport sector in Sweden is especially well suited as the ETS does not cover it, and it has a large carbon tax coverage. I construct my counterfactual by creating a synthetic version of Sweden through synthetic control method. Since the carbon tax pertains to the entire transportation sector, I do not have an easily obtainable counterfactor for a differences-in-differences (DiD) regression. In simple terms, the synthetic control method allows me to weigh other EU countries that do not have a carbon tax to create synthetic Sweden.

The remainder of this paper is structured as follows. In section two, I explain the theoretical anchoring behind carbon pricing and its effect on innovation, which role carbon taxes play concerning the ESR. I explain why Sweden and the transport sector are especially well suited to investigate the effect of carbon taxes on innovation. I also provide a critical synthesis and develop the main hypothesis. Section three explains why I chose the synthetic control method as my analysis method of choice and provides the synthetic control method's formal specifications. Moreover, I present my data sources and the steps conducted to arrive at the final dataset that I used for the analysis. Section four comprises a description of my results and robustness checks. Section five thoroughly discusses the results, deducts relevant political implications, links the results back to the economic theory, and explains that the results are generalizable to other countries outside Sweden. Finally, chapter six summarizes the advancements and points out future research opportunities.

2. Carbon Pricing and Innovation

The internalization of negative externalities, like carbon, leads companies to innovate to reduce the per unit cost of the externality they face (Acemoglu et al., 2012; Hicks, 1932; Pigou, 1920; Porter, 1991). In Europe, the Effort Sharing Regulation (ESR) sets binding annual GHG emission targets for its member states. A carbon tax can be used by these states to reduce GHG emissions through innovation in clean technology. Sweden plays a special role in carbon pricing as it was one of the first countries globally to implement a carbon tax at a high rate (Appendix 1). In this chapter, I will thoroughly examine each of the aforementioned components, explain their interconnectedness, embed them in the existing literature, and derive my hypothesis.

2.1. Economic theory on Carbon Pricing and Innovation

In a theoretically perfect market, sellers and buyers communicate effortlessly, and the market equilibrium equates to the producer's marginal cost and the consumer's willingness to pay. In reality, however, market failures, which are characterized by the inefficient distribution of products on the free market, often arise due to externalities. Externalities cause market failure as these are discommodities that a rational agent is incentivized to disown and avoid. Hence, the rules of market operations apply to discommodities but in reverse (Tybout, 1972). An environmental externality arises when the environmental damage of a good is not priced in the product. Hence, market outcomes are inefficient when consumers or companies are not exposed to the actual social cost attributed to their behavior (Knittel & Sandler, 2018). One solution to correct the distortion caused by negative externalities is introducing a Pigouvian tax so that the agents internalize the cost of their behavior (Pigou, 1920). Through the internalization of externalities, the value of the damage is factored into the actions that companies and consumers take. Hence, an efficient price level of the commodity and, importantly, an efficient level of emissions are reached (Lilliestam et al., 2020). The Pigouvian tax is often not applied to the actual externality but rather to the product most closely correlated with the externality. In the transportation sector, for instance, the carbon tax is applied to the average fossil carbon content of fuels because it is practically infeasible to tax the actual carbon emitted during usage.

An optimal Pigouvian tax incentivizes lower-cost abatement options to the level that matches the costs induced by the tax itself. In the automotive industry for example, abatement options can come in different forms, such as switching the fuel type of a vehicle, improving the fuel usage of internal combustion engines (ICEs), using a hybrid system, or entirely replacing the ICE with an electric motor (OECD Environment Directorate & International Energy Agency, 2001). The CO_2 abatement options that are patented fall under the Y02T category. The Y02T category refers to clean technologies related to transportation and is used as a critical variable for the analysis conducted in this paper.

Next to cost-minimization, a carbon tax generates incentives for developing and disseminating technologies that are less emitting than the prevailing standard. The assumption is that companies are motivated to innovate because they face a larger price for their emissions, increasing their production costs. Hence, companies are incentivized to invest strategically to reduce the ratio of emissions per production unit. Hicks (1932) was the first to make this assumption, which he coined the microeconomic-induced innovation hypothesis (IIH). He states that an increase in the price of input factors of production motivates invention. Porter (1991) and Acemoglu et al. (2012) expand on Hicks' microeconomic hypothesis and apply it to environmental policy. They hypothesize that a significant part of investments will flow to the development and commercialization of clean technologies as this is more economical than the cost incurred through continuing business as usual (Jaffe et al., 2003; Stavins, 2007). From a strategic point of view, companies that innovate early, so-called first movers, can take advantage of learning curve effects and patenting to attain a sustained competitive advantage compared to companies that do so later. In general terms, the "weak" Porter hypothesis (PH) asserts that stricter environmental policy regulations[1] stimulate innovation. It has to be differentiated from the "strong" Porter hypothesis, which argues that stricter regulations positively impact business performance. In this paper, I follow the rationale of the weak PH as a considerable strand of literature supports the first PH, while there is mixed empirical evidence concerning the second PH (Ambec et al., 2013; Palmer et al., 1995).

Carbon pricing, in the form of a carbon tax, can positively impact eco-innovation through the higher costs exerted on companies, government revenue allocated to carbon mitigation technology development, and a credible policy commitment. Carbon pricing is widely considered to be an economically viable option to inducing clean innovation as well as reducing GHG from a theoretical perspective (Baumol & Oates, 1988; Stavins, 2007), which is corroborated by empirical studies (Andersson, 2019; Elgie & McClay, 2013; Rivers & Schaufele, 2015).

2.2. The Effort Sharing Regulation and Carbon Taxes

According to the ESR, each EU member state has binding annual GHG emission targets for 2021-2030 regarding those sectors that the EU ETS does not cover. These sectors are transportation, buildings, agriculture, small industry (non-ETS), and waste. They account for around 60% of GHG emissions. Member states have different capacities to take action to reduce GHG emissions. Therefore, differentiating targets are allocated across the members according to the gross domestic product (GDP) per capita. Slight adjustments are made for countries with extraordinarily high

[1] This paper uses a credible carbon tax as a proxy for strict policy regulation. Although there is no clearcut definition for a credible carbon tax, I refer to a carbon tax that is implemented at a rate similar to Nordic European countries (Appendix 1).

GDP per capita so that these do not face excessive mitigation costs. Country-specific targets for 2030 range from 0% to 40% emission reduction compared to 2005 levels, while the legislation currently in place aims to reduce emissions by 30% across the entire EU (European Commission, 2022a). Although Iceland and Norway are not EU members, they committed themselves to being part of the ESR. According to the Unify Program (2022), which is funded by the LIFE program of the European Union, the ESR targets should be further increased if the EU wants to comply with the obligations under the Paris Agreement. Emissions should be reduced by at least 50% instead of the current 30% compared to 2005 (European Environment Agency, 2021).

In order to achieve current or even more ambitious targets while maintaining economic competitiveness, the green growth strategy is a common policy approach chosen. The aim of the green growth strategy, as laid out by the European Union's Green Deal, is to foster economic growth and development while decreasing GHG emissions. This decoupling of economic growth and environmental pollution is envisaged to be achieved through the development of clean technologies. According to Howard and Sylvan (2015) and Commission (2017), the most economical way to decrease the risks of climate change and foster innovation is to implement a carbon tax or an ETS. However, public support outside of academia is lower. Many politicians for instance believe that the effect on innovation and the environment are limited. Support increases when evidence is presented that carbon taxes indeed foster innovation and GHG mitigation (Andersson, 2019; Murray & Rivers, 2015). To date, only 17 out of the 30 EU ETS member states have implemented a carbon tax. Out of these 17 countries, predominantly Nordic countries implemented carbon taxes with a significant price level (Appendix 1). Correct empirical estimations of the effect of carbon taxes on eco-innovation and related GHG mitigation potential are crucial to foster political support and to ensure credible policy commitments.

The first wave of countries that implemented a carbon tax predominantly did so because of green governments and comprises of the countries Finland (1990), Norway (1990), Sweden (1990), Denmark (1992), Slovenia (1996), Estonia (2000), and Latvia (2004), and Liechtenstein (2008).

In order to mitigate GHG emissions not explicitly covered by the EU ETS and comply with the ESR, several other EU countries introduced carbon taxation after the introduction of the Effort Sharing Decision in 2008. This second wave of countries comprises Iceland (2010), Ireland (2010), Ukraine (2011), France (2014), Spain (2014), Portugal (2015), Luxembourg (2021), Netherlands (2021), and Germany (2021). The Effort Sharing Decision was introduced in 2008 and set national emission targets for 2013 to 2020. The Effort Sharing Decision then transitioned into the Effort Sharing Regulation, which sets targets for 2021 until 2030. The national carbon taxes vary in GHG emission coverage, rate, and percentage of GHG emissions overlapping with the EU ETS (World Bank, 2022).

The transport sector accounted for around 36% of ESR emissions in 2019. Among all the ESR sectors, transport has the highest intended reduction until 2030. Nevertheless, between 2005 and 2019, the total reductions of the transport sector comprised only 5% of the reductions achieved in the total ESR, corresponding to 13 Mt CO_{2e} (Unify Program, 2022). As emission abatement in the transportation sector poses a significant challenge, an increasing number of countries have implemented a carbon taxation system, particularly for this sector. On the EU level, the European Commission proposed the "Fit for 55" legislative package in July 2021. This package proposes an ETS that also covers road transport, which would make it the largest ETS to apply to road transport. It is intended to exist separate from the EU ETS and regulated fuel suppliers, which will be responsible for incorporating the carbon cost. Applying carbon pricing to the road transport sector increases the price level of fuel, which according to Hicks (1932) and Porter (1991) and Acemoglu et al. (2012), increases innovation in clean technologies. There is an incentive to reduce the CO_{2e} content per liter of fuel to face lower taxation. CO_{2e} reduction might be accomplished by increasing vehicles' fuel efficiency or substituting conventional fuels with alternative fuels or energy sources such as electric batteries. The increase in the fuel price is ultimately passed to the consumer, who will strive to save fuel by buying increasingly environmentally friendly vehicles. Producers can capitalize on this trend by investing in the development of vehicles with low carbon emissions, which customers prefer (Aghion et al., 2016).

In contrast to an ETS, the carbon tax price is less volatile, which allows the risk-averse investor to make more confident investment decisions. Thus, firms can make significant clean technology-related investments (International Energy Agency, 2007). Analyzing the Swedish carbon tax's effect on innovation in the transportation sector may provide insights that can be used to fine-tune the implementation of the "Fit for 55" legislative package. Moreover, Sweden, as a pioneer in the early and credible introduction of a carbon tax, has a vital role in showing other countries that carbon taxes characterized by high price levels allow for innovation and GHG mitigation in harmony with economic growth. This paper fills the gap in ex-post empirical studies on the causal effect of carbon taxes on eco-innovation. I provide an empirical analysis of the effect of introducing a carbon tax in Sweden on eco-innovation. Eco-innovation allows for long-term GHG mitigation. Given the lack of available ex-post studies on the effect of carbon taxes on eco-innovation, the findings of this paper aim to corroborate confidence in implementing carbon taxes. Politicians implementing less efficient long-term mitigation measures will face challenges in reaching current targets set under the Paris Climate Agreement.

2.3. Sweden and Carbon Taxes

In Sweden, the Social Democrats were the first to recognize the threat of climate change and suggest a tax. In 1990[2]

[2] I use 1990 as the year of my policy intervention. Other empirical studies on the carbon tax in Sweden use either 1990 or 1991 as the year of policy

this tax was promulgated by the Social Democratic government (Collier & Löfstedt, 1997). Sweden has a long history of taxing energy products to raise tax revenue and has been taxing petrol since 1924, diesel since 1937, and coal, oil, and electricity for heating purposes since the 1950s. This preexisting infrastructure for taxing energy products paved the way for implementing a carbon tax in Sweden (Jonsson et al., 2020). The carbon tax remains the fundament of Swedish climate policy today (Ministry of Finance, 2021). Hammar and Åkerfeldt (2011) describe the significant tax reform in 1990–1991 as "grön skatteväxling" translating to a "green tax shift" as other taxes, such as labor taxes, and energy taxes were reduced simultaneously to encourage green growth (Regeringskansliet, 2014). The marginal personal income tax rate was reduced from the highest rate of 80% to 50%, and the corporate tax rate from 57% to 30% (Jonsson et al., 2020). Nevertheless, Sweden also broadened the coverage of its value-added tax (VAT) in 1990 to pertain to gasoline and diesel. A VAT of 25% is applied to transport fuel, exercise taxes, and producer margin (Andersson, 2019). Since implementing the carbon tax more than 30 years ago, Sweden has achieved green growth because it reduced GHG emissions while maintaining GDP growth. GDP per capita increased by over 50% between 1990 and 2021 in real terms (OECD, 2022).

Swedish carbon tax revenues comprise around 1% of the government's total tax revenues, corresponding to SEK 22.2 billion ($2.3 billion) (Natur Vårds Verket, 2019). Although Sweden does not use the carbon tax revenues for direct green spending, which is revenue earmarked for climate protection, they use 50% of the revenues as general funds, which go to the government budget, and the other 50% for revenue recycling. Revenue recycling refers to income redistribution to firms and consumers through tax reductions or subsidies (Lilliestam et al., 2020). From 1990 until 2004, the revenue increased, stabilized until 2010, and decreased slightly over the last decade. As fewer fuels or fuels with lower GHG emissions are used, fewer tax revenues are collected, which is intended by the system's design. The carbon tax revenue collected now comprises 95% of taxes on motor fuels. However, heating fuels made up a large percentage of the collected tax revenue when first implemented. Since 1990 fossil heating fuels have been phased out, and their usage has decreased by 85% and now represents only 2% of Sweden's total GHG emissions. Sweden replaced fossil fuel heating with district heating and heat pumps, a more sustainable and holistic system (Ministry of Finance, 2021).

Although combating climate change had extensive political support in the period from 1980 until 2000, concerns about carbon leakage and competitiveness in a global economy led to the industry paying only 25% of the full rate and exemptions for the electricity industry. Between 1993 and 2015, the tax rate for the Swedish industry varied between 21% and 50% of the full rate and was gradually phased out with the introduction of the EU ETS (OECD, 2016) (Appendix 2). Therefore, the carbon tax had a relatively low impact on the industry (Johansson, 2000). Nevertheless, because tax rates differ across energy products and users, other sectors, such as residential, commercial, or road transport, are affected more significantly (Appendix 3). The current carbon tax in Sweden, with a rate of SEK 1200 per metric ton of CO_2, is the highest in the world (International Energy Agency, 2022).

The introduction of the carbon tax resulted in a low administration as the tax is levied on importers, distributors, and large consumers rather than large numbers of final consumers. Gasoline, for example, is already taxed at the point of import or wholesale, meaning that neither the gas station operator nor the final customer is taxed directly. The legal incidences differ from the economic ones as the tax is administered to importers, distributors, and large consumers, but the economic costs are passed down to the final consumer.

Next to the industry sector, which emits 60,176 t CO_2, road transportation is the second largest emitting sector of CO_2 in Sweden with 21,241 t CO_2 (OECD, 2016). Carbon taxes pertain to 91% of the emissions emitted by the road transportation sector (OECD, 2016). Sweden has a material interest in reducing CO_2 emissions in the transport sector as the automotive industry is its largest export sector with a value of around € 11B annually and employs the highest number of people of all industries in absolute numbers (OEC, 2022). Although Sweden is a relatively small EU country, with a population of only 10.35 m as of 2022, it houses the headquarter of the large truck manufacturers Scania CV AB and the Volvo Group, and the personal vehicle manufacturer Volvo Car AB. Sweden's strong economic position is based on its export-oriented industry. Foreign international competition constantly induces pressure for change, promoting innovation in Swedish firms.

Sweden is particularly well suited to study the effects of carbon taxation on innovation as Swedish companies have a strong focus on achieving growth in line with the government's policies. This is mainly done through product improvement and innovation by investing in R&D (Johansson, 2000). Sweden has the second largest Business R&D intensity of all countries in the Organization for Economic Cooperation and Development (OECD) (Figure 1). Figure 1 shows the Business enterprise expenditure on R&D (BERD) adjusted for industrial structure, which measures a country's business R&D intensity assuming it had an OECD average industrial structure. BERD represents the components of the Government expenditure on R&D (GERD) incurred by units belonging to the Business enterprise sector. The unit of measurement is the BERD as a percentage of gross value added (GVA) in industry (OECD, 2020).

According to the European Innovation Scoreboard (2022b), Sweden is an innovation leader with a performance of 135.7% of the EU average. Sweden scores especially high in public-private co-publications (381.4%), international scientific co-publications (241.1%), intellectual assets such as PCT

intervention. The actual tax affected consumers in 1991. However, antecedent effects due to press coverage on the promulgation might already exist in 1990.

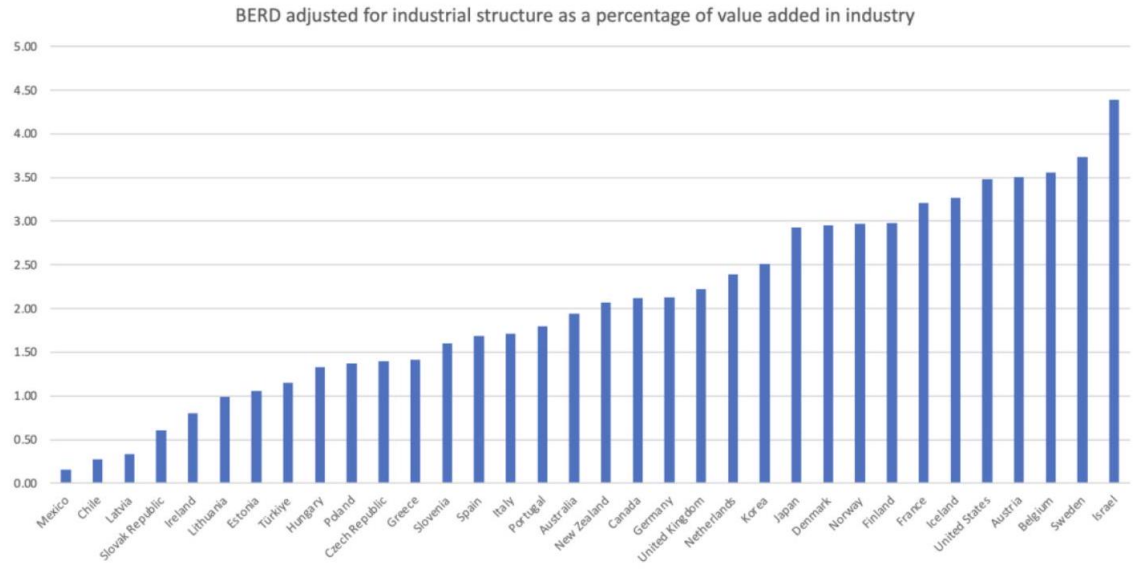


Figure 1: BERD as a percentage of value added in industry (OECD, 2020)

patent applications (150.6%), and eco-innovations (123%). Sweden is among the countries that most significantly invest in R&D in relation to their GDP. Sweden's total R&D investments - an important input for innovation - amount to around 4% of GDP in 2020 (SCB, 2020). The Swedish innovation system comprises an integrated public and business sector, whereby the business sector contributes to nearly 72% of all expenditures.

The configuration of business and public innovation structures is especially pronounced in the automotive industry. The "Fordonsstrategisk forskning och Innovation" (FFI), which translates to Strategic Vehicle Research and Innovation Programme, is the largest collaboration between the Swedish state and the automotive industry. Vinnova, the government's innovation agency, collaborates with the automotive industry (Scania CV AB, AB Volvo, Volvo Car Group, and FKG). From 2009 until 2020, the FFI conducted around 900 research projects with over 500 Ph.D. researchers worth € 100 million annually, with half of the funding coming from the government (VINNOVA SE, 2022). Many of those projects concern reducing energy use per vehicle, linked to clean innovation.

While the Swedish government does not have a mandate to spend the revenues earned from the carbon tax on green *R&D* projects, it still uses some of this revenue to fund research projects concerning clean innovation in transportation. This may be because the Swedish government deems environmentally friendlier vehicles important for the sector's international competitiveness. In Sweden, the automotive industry comprises 14% of its economy. It is not uncommon for governments to participate in the research of climate-friendly technologies after introducing a carbon tax. In Sweden, the government started to invest in the research of clean transportation technologies in the early 2000s and increased funding with the introduction of the FFI (VINNOVA SE, 2022). In my study, I treat the government's research projects as part of the effects of a carbon tax on innovation.

The implementation of the carbon tax was especially successful in reducing GHG emissions in the transport and heating sector (Natur Vårds Verket, 2022a). According to the Swedish environmental protection agency, GHG emissions have been reduced by 32% since 1990, especially during the last 20 years (Natur Vårds Verket, 2022b). Although GHG emissions have been reduced, Sweden's carbon tax is subject to criticism. Experts from the Stockholm School of Economics criticize the carbon tax because it is likely that the effect on emissions reduction is below the effect that could have achieved (Jonsson et al., 2020). While many small companies cut their emissions significantly, the largest polluters, manufacturers of steel and concrete, increased emissions. The failure in nudging large polluters to emit less GHG emissions is due to the cap, which imposes a maximum tax a company must pay. The government's idea was to limit the competitive threat a carbon tax might pose to large steel and cement producers on an international stage, as these sectors were considered of national interest due to extensive lobbying (Karakaya et al., 2018). While their competitiveness was preserved, large polluters were exempt from paying for marginal emissions beyond the cap until they were covered by the EU ETS (Lundberg, 2014).

2.4. Critical Synthesis

In my literature review, I focus on those studies that evaluate the primary aim of green policies, especially the carbon tax policy, under the Paris Climate Agreement in 2015. The Paris Climate Agreement aims to limit GHG emissions to well below 2°C compared to pre-industrial levels and to pursue efforts to limit global warming even further to 1.5°C. In contrast to the Kyoto Protocol, which focuses rather on

ephemeral emission reduction, the Paris Agreement is about complete decarbonization and takes a long-term perspective. I emphasize the dissimilarity between emission reduction and entire elimination as differing analysis methods are deployed for policy evaluation (Patt & Lilliestam, 2018). My paper focuses on the carbon tax's long-term effects, which aligns with the efforts under the Paris Climate Agreement. I am interested in the effects of a carbon tax on a shift towards clean innovation and take a dynamic perspective assessing effects over decades, which is the time horizon carbon taxes are designed to operate over (Fuss et al., 2018). I review those ex-post studies that take the cost-effectiveness framework, focusing on deep technological innovation in long-lived carbon-emitting capital stocks such as transportation.

Moreover, I mainly examine peer-reviewed papers, except for Moore et al.'s (2021) paper, which is a working paper but is the only other research I could find which also studies the introduction of the carbon tax in Sweden and its effect on the transportation sector. I perform my search analysis by using the academic search engines ScienceDirect and Google Scholar and searching for terms such as ("carbon pricing", "carbon tax Sweden", "carbon tax") for my research domain. I use more general search terms for literature on carbon pricing theory.

Concerning my theory section in chapter 2.1, an extensive number of macro as well as microeconomic studies exist concerning the IIH (Newell et al., 1999; Popp, 2002). Regarding carbon pricing and innovation, most empirical evidence supports the first Porter hypothesis. In an environmental context, Goulder and Schneider (1999), Gerlagh (2008), and Acemoglu et al. (2012) find empirical support for the innovation effect, namely that strict environmental regulations trigger the introduction of clean technologies, which makes production more efficient. Brunnermeier and Cohen (2003) and Lee et al. (2011) explicitly focus their research on the weak Porter hypothesis and find that environmental policies positively affect clean innovation in the auto industry.

Most of the research on carbon taxes focuses on Nordic countries, as these are among the world's oldest and highest-priced carbon taxes (Lilliestam et al., 2020). Nevertheless, there are also some studies concerning the carbon tax in British Columbia, which was implemented in 2008 and is one of the few examples of a high carbon tax outside of Europe.

Bohlin (1998) investigates the effect of the carbon tax in Sweden on CO_2 emissions in different sectors from 1990 until 1995. He finds a significant effect of carbon taxes on CO_2 savings in the district heating sector and no effect in the transportation sector. Bohlin's study only covers the first five years since the carbon tax's introduction and cannot observe the long-run effects of the policy implementation.

More than ten years later, Lin and Li (2011) investigate the effects of carbon taxes on GHG emissions in Sweden, Denmark, Finland, the Netherlands, and Norway between 1981 and 2008, while the period from 1981 until 1989 is the pre-treatment period. They use a DiD model and only find a significant effect for Finland. They estimate a 1.7% reduction in emission increase compared to the hypothetical scenario without a carbon tax implementation. Lin and Li (2011) point out that tax exemptions for energy-intensive and manufacturing industries are likely the reason they found no emissions growth reduction for the other countries analyzed.

The most prominent paper regarding carbon taxes in British Columbia concerns the impact on gasoline demand (Rivers & Schaufele, 2015). They analyze the period from 2007 until 2011 and find that the carbon tax has a larger negative effect on gasoline consumption than other taxes. They conclude that this can be attributed to the salience of carbon taxes in public disputation and the press.

Shmelev and Speck (2018) analyze the effect of the introduction of the carbon tax in Sweden on CO_2 emissions from 1961 until 2012. The years before 1990 are again used as a pre-treatment period. They find that the carbon tax lowered the CO_2 emissions from the use of petrol, but not so for other energy sources studied. They find several coefficients for the effect on petrol as they implement a multitude of econometric models.

Recent developments concerning econometric policy evaluation tools have made it possible to overcome common issues related to the DiD method. Andersson (2019) applied the synthetic control method to analyze the effect of introducing a carbon tax on emissions from the transport sector from 1990 until 2005, with a pre-treatment period starting in 1960. His donor pool consists of a group of OECD countries that did not implement a carbon tax during the period studied. He finds that the carbon tax reduced emissions in the transport sector by 6% compared to synthetic

Sweden's scenario without introducing a carbon tax. Not only do his findings oppose Lin and Li's (2011), but he also scrutinizes their research design. He mentions two main points of criticism. First, Lin and Li (2011) use total CO_2 emissions as the dependent variable. Therefore, they combine treated and untreated sectors, although all the countries they analyze have sectors of the economy that are exempt from carbon taxes. Second, they include covariates in their DiD model that are related to their outcome variable. While this is allowed in the synthetic control method, it biases results when the DiD method is used.

The only peer-reviewed article that is an ex-post empirical assessment of a carbon tax and innovation is the study by Cheng et al. (2021). They study the dynamics between carbon tax revenue and energy innovation in Sweden from 1990 until 2019. They deploy the recently developed Quantile-on-Quantile Regression framework to observe whether there is a linear relationship between tax revenue and innovation. They observe that when there is a low carbon tax revenue, a higher penetration of energy innovation is desired. They state that this might be because a low carbon tax burden might be less costly for a firm compared to the costs of implementing energy innovation. This effect disappears for larger carbon tax revenues.

Even though Moore et al.'s (2021) paper is a working paper, I choose to include it in this peer review because it has a similar research domain to my paper, as it also investigates the carbon tax and Sweden's transport sector. They use the

synthetic control method and find a positive effect of carbon taxes on clean innovation, with an average increase in clean patents of 7.37 per year from 1990 until 1999. Sweden started to finance R&D programs to boost clean innovation beginning in 1999. They see the government's financing of R&D programs as policy interventions unconnected to the carbon tax and hence as a confounding variable they cannot control. Therefore, they decide not to investigate the period after 1999 further. In my study, I take a similar approach to Cheng et al. (2021) and Andersson (2019) and do not regard those government initiatives as separate from the carbon tax. Like Andersson (2019), they construct synthetic Sweden using OECD countries as their donor pool. Using a donor pool that comprises OECD countries is a sensible approach for Andersson's (2019) study, which has GHG emissions as a dependent variable. Using the same donor pool when having patents as a dependent variable might introduce some issues. Using countries like the United States or Japan in the donor pool to construct a counterfactual might bias results as those countries' innovation infrastructure is significantly larger than that of European countries. I provide a more detailed review of Moore et al.'s (2021) study and a comparison of results in the discussion section.

Although Calel and Dechezleprêtre's (2016) study does not concern carbon taxes but the ETS, I chose to include it in this literature review as it is considered one of the most prominent studies on carbon pricing and innovation. It analyzes the EU ETS' effect on directed technological change by using patent count as a proxy for innovation, which is regarded as one of the most robust indicators of innovation (Hagedoorn & Cloodt, 2003). Furthermore, causal claims are made using the DiD approach that allows for a precise interpretation of the estimated results (Teixidó et al., 2019). Calel and Dechezleprêtre (2016) find that the share of lowcarbon patents among the companies regulated by the EU ETS rises significantly during the first five years after the launch of the EU ETS. Notably, they find no such phenomenon concerning non-regulated firms. They estimate that the share of low-carbon technology patents increases by 36.2% compared to a scenario without the EU ETS, while regulated firms continue their patenting behavior for other technologies. One limitation is that the effect could have also been driven by confounding variables such as a rapidly rising oil price, which has not been investigated closer.

Regarding the transport sector, empirical studies find that there is a positive effect of fuel prices on innovation in clean technologies concerning automotives such as improvements in the fuel efficiency of ICE, alternative fuels for the ICE, hybrid vehicles, or electric vehicles (Aghion et al., 2016; Crabb & Johnson, 2010; Hascic et al., 2009). These studies estimate the potential effect of carbon taxation on innovation in the transport sector. A fuel tax is similar to a carbon tax levied on fuel, but the GHG emissions of the fuel are not explicitly accounted for in a general fuel tax. Moreover, the carbon tax levied in Sweden is larger than the fuel tax levied in many other countries (Andersson, 2019; Rivers & Schaufele, 2015).

The literature on carbon taxes and innovation remains scarce. Previous literature predominantly focuses on the environmental effects of introducing a carbon tax. Advancements in econometric models in recent years allowed researchers to apply new methods to investigate the effect of the carbon tax on innovation (Cheng et al., 2021; Moore et al., 2021). It has not yet been sufficiently clarified what the effect of the introduction of a carbon tax is on innovation, particularly regarding recent terms. Although Cheng et al. (2021) do investigate the relationship between a carbon tax and innovation until 2019, their findings lack sufficient interpretation and contextualization as they use complex econometric models for their research which are challenging to interpret. Moreover, their research rather concerns the relationship between carbon taxes and innovation rather than causal dependencies.

2.5. Hypothesis Development

I now develop my main hypothesis by combining my findings on the theory and literature review on carbon pricing. Pigou (1920) suggests that introducing a Pigouvian tax, such as the carbon tax, leads agents to internalize the cost of their behavior. Once the negative externalities are priced in the cost of the product, companies are incentivized to innovate their production methods to reduce the per unit cost of the negative externality they are now facing (Hicks, 1932). More applicable to carbon taxes, Porter (1991) and Acemoglu et al. (2012) apply Hick's (1932) hypothesis to climate policy. They hypothesize that climate policies lead to clean technology development as this is economically more attractive than the cost associated with abatement. Previous literature on carbon taxes and innovation indeed explores the economic relationship between the cost of abatement and the cost to innovate (Cheng et al., 2021). However, it lacks the identification of the overall effect of clean innovation. The only other study on carbon taxation and clean innovation merely investigates a period of roughly ten years and finds a moderate but positive effect on innovation (Moore et al., 2021).

In my paper, I investigate the effect of the introduction of a carbon tax on clean innovation over a prolonged period. It is unfeasible to study the effect of a carbon tax across several countries, given different sector coverages and introduction periods. Therefore, I study the effect in Sweden and discuss to which degree my findings are generalizable to other countries.

Based on this, I hypothesize that a positive relationship exists between introducing a carbon tax and clean innovation in the long run.

I test this hypothesis by looking at Sweden's transport sector and a treatment period that stretches from 1990 until 2018. I chose to focus on Sweden as it was not only one of the first countries to introduce a carbon tax but also has the largest policy tax rate (Appendix 1). The transport sector in Sweden is well suited as the carbon tax covers 91% of its emissions, while the EU ETS covers 0% (Appendix 2). Sweden has a sizeable automotive industry compared to other Nordic countries, so implementing innovation to reduce GHG

in this sector has exceptionally high materiality for Sweden (Atradius, 2019). I use the synthetic control method to measure the effect of the carbon tax on clean innovation. To construct my synthetic Sweden, I use the EU ETS member states which do not have a carbon tax as my donor pool.

3. A Methodological Approach to Estimate the Effect of a Policy Introduction on Innovation at the Country Level

To overcome issues concerning the DiD method, I use the synthetic control method to determine the effect of the carbon tax implementation on clean innovation. This chapter explains why the synthetic control method is well-suited to answer my hypothesis. Next, the formal aspects of the synthetic control method are presented. Finally, I explain the data I use to construct my synthetic control and where it originates.

3.1. The Empirical Analysis Method

Previous studies evaluating the effect of a policy on innovation at the company level have mainly used methods such as the DiD method to compute a causal effect (see e.g., the study of Calel and Dechezleprêtre (2016)). The DID method is applied to longitudinal data for which a treatment and control group exist. The effect of an intervention is estimated by measuring the changes in outcomes over time between the treatment and control group. In the case of Calel and Dechezleprêtre (2016) the units exposed are the ETS-regulated companies, and those not covered by the system form the control group. Contrary to the ETS system, which does not affect all firms in a country's sector, the carbon tax usually affects all companies in the same sector. Hence, no set of units can be easily chosen to form the counterfactual. By deploying the synthetic control method, I can create a counterfactual by using other countries in the EU that do not have a carbon tax. In the synthetic control method, the untreated units are weighted so that they mimic the behavior of the treatment unit as precisely as possible without actually being treated themselves. The weighting of the comparison units, which form the donor pool, is done by matching their pre-exposure trends based on predictor variables to estimate the counterfactual optimally. The then-created counterfactual is called the synthetic control unit.

The difference between the evolution of the treatment unit, Sweden, and the synthetic control unit is the gap that represents the effect of the policy intervention. The large portion of existing literature on carbon pricing and innovation does not use a quasi-experimental design because of the lack of a counterfactual. Through the deployment of the synthetic control method, comparative case studies, which were previously not feasible, it has now become possible at the country level. This method was introduced in 2003 and is regarded as one of the most significant innovations in policy evaluation (Abadie & Gardeazabal, 2003). More recent advancements (Abadie et al., 2010, 2014) allow it to be a powerful generalize of the DID approach (Cunningham, 2021). The gap between the scenario in which the government policy would not have been introduced and the current scenario, which is the introduction of the policy, can be estimated through the synthetic control method. The synthetic control method is increasingly used in academia and the industry, especially the tech industry, as it is easy to interpret and can deal with large-scale settings. Recently Andersson (2019) used the synthetic control method to determine the impact of carbon taxes and (VAT) on transport fuel on GHG emissions in Sweden.

Abadie (2021) argues that the synthetic control method poses advantages over common applied econometric regression-based methods. Abadie (2021) points out that a regression-based approach can be useful for studying the short-term effects of a policy introduction where it is estimated that the effect has a significant magnitude. Nevertheless, time-series techniques lose explanatory power for estimating medium and long-term effects due to the presence of confounding variables that pose a shock to the outcome of interest.

In conducting a DID regression, the researcher must make a parallel trends assumption before the intervention to control for selection effects by accounting for time-fixed and unit-fixed effects. In analyzing a policy for a specific country like Sweden, the synthetic control method has the advantage that no parallel trends assumption is needed. The underlying idea of the synthetic control method is to exploit the temporal variation in the data in contrast to the cross-sectional one. Abadie (2021) explores the technical advantages of the synthetic control method over regression-based methods in detail. His three main arguments are that no extrapolation is conducted, there is transparency to the fit, and transparency to the counterfactual.

Given the advantages the synthetic control method poses, I apply the synthetic control method to estimate the effect of the introduction of the carbon tax in Sweden in 1990 on clean technology innovation in the transport sector.

3.2. Formal Aspects of the Synthetic Control Method

I follow a similar approach to Abadie (2021) to construct synthetic Sweden. Abadie (2021) provides an exhaustive explanation of the formal aspects of the synthetic control method. In this section, I delineate the key aspects that are needed to understand how to synthetic control method works. Hence, I retrieved data for J+1 units: $j = 1, 2, \ldots, JJ+1$. My first unit, $j = 1$, represents the treatment unit Sweden, which is affected by the carbon tax introduction. The other units represent the donor pool, $j = 2, \ldots, J+1$, which are the units unaffected by the carbon tax. The entire data span, 1985 to 2018, are T periods. T_0 represents the periods before the intervention. For each time period, t, and unit, j, the outcome Y_{jt} is observed. For every unit, j, I have a set of k predictors of the outcome, $X_{1j}, \ldots, X_{kj}$. A common approach is to include the outcome variable itself as one of the predictors. I follow this approach and include the outcome variable in the set of k predictors. The outcome variable itself is not affected by the treatment before the intervention. For the treatment unit, $j = 1$, during the period, $t > T_0$, Y^I_{jt} denotes the outcome variable with the policy intervention. The

outcome variable without the policy intervention is denoted as Y_{jt}^N. The effect of the policy intervention on the treatment unit in the period, $t > T_0$, is:

$$\tau_{1t} = Y_{jt}^I - Y_{jt}^N \tag{1}$$

The synthetic control unit is the weighted average of the units in the donor pool that most closely match the treatment unit. Hence, the synthetic control unit is a $J \times 1$ vector of the weights, $W = (w_2, \ldots, w_{J+1})'$. The synthetic control estimator of the outcome variable is:

$$\hat{Y}_{lt}^N = \sum_{j=2}^{J+1} w_j Y_{jt} \tag{2}$$

To avoid extrapolation, Abadie (2021) uses the approach to restrict the weights to be positive and the sum being one. Restricting the weights to be nonnegative has the advantage of not having to use regression-based methods to extrapolate. This makes the result more transparent and easier to interpret.

I follow Abadie and Gardezabal's (2003) and Abadie et al.'s (2010) proposal to choose the weights that result in the synthetic control unit being the closest resemblance to the pre-intervention values concerning the treated unit of predictors of the outcome variable. The $k \times 1$ vectors $X_1, \ldots, X_{J+1}$ represent the values of the predictors for the units $j = 1, \ldots, J+1$. The $k \times J$ matrix, $X_0 = [X_2 \cdots X_{J+1}]$, represents the values of predictors for the untreated units, J. Hence, I chose the weights such that equation 3 is minimized while the weights are non-negative and sum up to one.

$$\|X_1 - X_0 W\| \tag{3}$$

In simple terms, the resulting synthetic control unit is a weighted average of untreated units that minimize the difference to the control unit concerning key predictors of the outcome variable. W(V) are the weights that are assigned to the untreated units. W is a function $V = (v_1, \ldots, v_k)$, representing the vector of predictor weights of k predictors. To solve equation 3 and choose the optimal values of V, I again follow Abadie and Gardezabal's (2003) and Abadie et al.'s (2010) approach and choose V so that W(V) minimizes the mean squared prediction error (MSPE) of the synthetic control unit.

Finally, the estimated effect of the intervention for the treated unit in period $t = T_0 + 1, \ldots, T$ is defined as

$$\hat{\tau}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j * Y_{jt} \tag{4}$$

3.3. Data

I use patent data from the European Patent Office's (EPO) Worldwide Patent Statistical Database (PATSTAT) as a proxy of innovation. Using patent data allows me to conduct a detailed analysis of innovation activity induced by the policy intervention. While there is consensus in academic research that patent data is one of the best proxies to measure innovation, there are certain limitations. One limitation is that not all inventions are ultimately patented. Some inventions, which for example, may have a relatively sizeable environmental impact, do not have enough economic possibilities. Hence, those innovations may not justify the cost of patenting (OECD, 2009). According to Pavitt (1988), Strategic considerations may lead the inventor to keep the invention secret to receive an alternative form of protection, such as protection under a trade secret, which results in the patent not appearing in the patent data. Another limitation is that the value distribution of patents is highly skewed (Harhoff et al., 1999). Thus, the number of patents issued does not directly translate into the actual value of the underlying patent. Some patents have considerable economic importance for corporations, while others have limited economic value.

Moreover, there are differences in patent law and practices from country to country, which limits the comparability of patent statistics across countries to some extent. To overcome this issue, I chose to use homogeneous patent data originating from the largest patent offices, such as the EPO and Patent Cooperation Treaty (PCT), rather than smaller national patent offices. If an inventor, for instance, files an international patent application under the PCT, the applicant simultaneously receives protection for the invention in a large number of countries. Using data from these large patent offices controls home bias, which refers to inventors being more probable to file a patent application at their local patent offices. According to Frietsch and Schmoch (2009), the home bias can be overcome by only analyzing patent families with one or more multinational filing at the EPO or PCT. By only including patent applications filed to the EPO or PCT, I overcome the limitation that some patent applications have a higher worth than others. When applying for patent applications at the major patent offices, inventors must pay substantially higher fees than when applying to domestic offices. Hence, selecting only those applications filed at the EPO or PCT discards those with a low expected commercial value.

I am able to distinguish clean technologies by sector by using the Y02 classification system, which was created by the EPO, the International Centre on Trade and Sustainable Development (ICTSD), and the United Nations Environmental Program (UNEP). The Y02 classification system is considered the most accurate representation of patents relating to climate change mitigation technologies available today (Calel & Dechezleprêtre, 2016). Calel and Dechezleprêtre (2016) conclude that it is becoming the international standard for studies on clean innovation. The Y02 scheme is a cross-scheme that overcomes the bias of finding multidisciplinary patents when searching for patent publications using technology fields. The Y02 scheme allows for retrieving patents belonging to several Cooperative Patent Classification (CPC) technology fields (Angelucci et al., 2018). Hence, I focus my analysis on counting the annual Y02 patent frequency as a proxy for directed technological change toward clean technology.

I use the most recent version of PATSTAT, the 2022 version

(European Patent Office, 2022), and retrieved Y02T patents. Y02T patents are Y02 patents concerning the transportation sector. In my PATSTAT query, I joined multiple tables to extract a final data frame that contains the application id of each patent, the inventor's country code, the inpadoc family id[3], the Y02T subclass, and earliest filing date. To record the invention's date, I use the earliest patent application date as it has been shown that this date closest resembles the actual innovation activity. My timeframe concerns all patent applications from 1958 until 2018. As the EPO and PCT were initiated in 1978, data stretches back earliest to 1978. However, for my synthetic control method, I can only use Y02T data from 1985 onwards as the predictors I retrieve from the OECD statistics database only date back to the earliest 1985.

To measure innovation in the transport sector, I use a similar methodological approach as Moore et al. (2021). First, I exclude all Y02T subcategories concerning aeronautics or air transport (Y02T 50/00) and Maritime or waterways transport (Y02T 70/00). An overview of the remaining Y02T subclasses and their respective frequency can be found in Appendix 4. The final data frame I retrieved from PATSTAT concerns all patent applications worldwide for my timeframe and applications of interest. It consists out of 526,456 observations and the variables *appln_id*, *person_id*, *person_ctry_code*, *indpadoc_family_id*, *cpc_class_symbol*, *and earliest_filing_date*.

Instead of counting every application id, I count patent families according to the inpadoc family id. The number of patent families in Sweden translates to the actual count of new technologies invented rather than the number of patent applications. Patent applications can have strongly differing values and might represent rather incremental improvements in the same technology rather than actual innovation. According to Harhoff et al. (2003), the size of the patent family is a good proxy for the value of the underlying technological invention. Therefore, I group all the observations which belong to one Y02T subclass while taking the sum of the number of family applications to get the size of the patent family. Many patent families have multiple inventors, which may be from differing countries. Hence, I weigh the inventor's country location within each family and finally compute a matrix that contains the weighted number of Y02T family applications for each country in the donor pool per year. It is important to note that the inventor must not be a natural person, and the inventor often is represented as a company's office in a certain location. Therefore, the location used is either the natural person's country of residence or the location where the company conducted the research. Each column of this matrix represents my outcome variable for each country, which I use for the synthetic control method.

Figure 2 illustrates the development of the number of Y02T patent families annually for Sweden from 1978 until 2014.

[3] The inpadoc family id concerns all patent application documents covering a certain technology (European Patent Office, 2022)

The last step I conduct before I can use the data frame for the synthetic control method is that I exclude countries affected by a similar policy as the carbon tax. My initial set of EU ETS member states consists of the 27 EU countries plus Iceland, Liechtenstein, and Norway. First, I exclude those countries that introduced a carbon tax: Denmark, Estonia, Finland, France, Iceland, Ireland, Latvia, Liechtenstein, Netherlands, Norway, Portugal, Slovenia, and Spain. Second, I exclude Germany and Italy because these countries implemented changes to their fuel tax. Albeit the United Kingdom currently is not a member state of the EU ETS, it was so during my period analyzed. I also exclude the United Kingdom as it also implemented a change to the fuel tax. Hence, my final dataset representing the donor countries consists of the 14 countries: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Greece, Hungary, Lithuania, Luxembourg, Malta, Poland, Romania, and Slovakia.

To predict my outcome variable, Y02T patents, I use the predictors listed in Table 1.

Total climate change mitigation patents refer to the total fraction of climate change mitigation-related patents by the investor's country of residence. This variable includes patents from every sector of the economy. Total triadic patents are those filed at every one of the three largest patent offices, namely the USPTO, JPO, and EPO. Transport (WIPO) IP5 patents are defined as patent families concerning transportation by the Word Intellectual Property Office (WIPO). They are filed in at least 2 of the five largest intellectual property offices. The five largest are the EPO, JPO, the Korean Intellectual Property Office (KIPO), the USPTO, and the State Intellectual Property Office of the People's Republic of China (CNIPA). The B60 patent class is defined by the WIPO and is similar to the transport (WIPO) predictor but refers to those patents that concern all vehicles except rail, maritime aircraft, or space-related vehicles. I include those B60 patents that are filed at the EPO. Including predictors that refer to patent applications at all three, at least two of the three, and only one of the three, I create robustness against possible selection bias.

The predictors total climate change mitigation patents, climate transport EPO, and Transport (WIPO) IP5 patent family measure how much knowledge a country has concerning climate change mitigation technologies and automobile-specific technologies (OECD, 2018a). Total climate change mitigation patents are those filed at the EPO; climate transport EPO are those patents that concern climate mitigation technology in the entire transport sector filed at the EPO. The larger the value of accumulated knowledge, the lower the cost to innovate in the same domain in the future. Hence, countries that invented clean technologies in the past will likely continue to do so in the future. The total triadic patent variable serves as a proxy of the country's innovative capacity. I follow Abadie's (2021) logic and include the outcome variable as a predictor variable, as this usually increases the model's predictive power.

Similar to Raghupathi and Raghupathi (2017), I include GDP and enrollment in tertiary education as economic in-

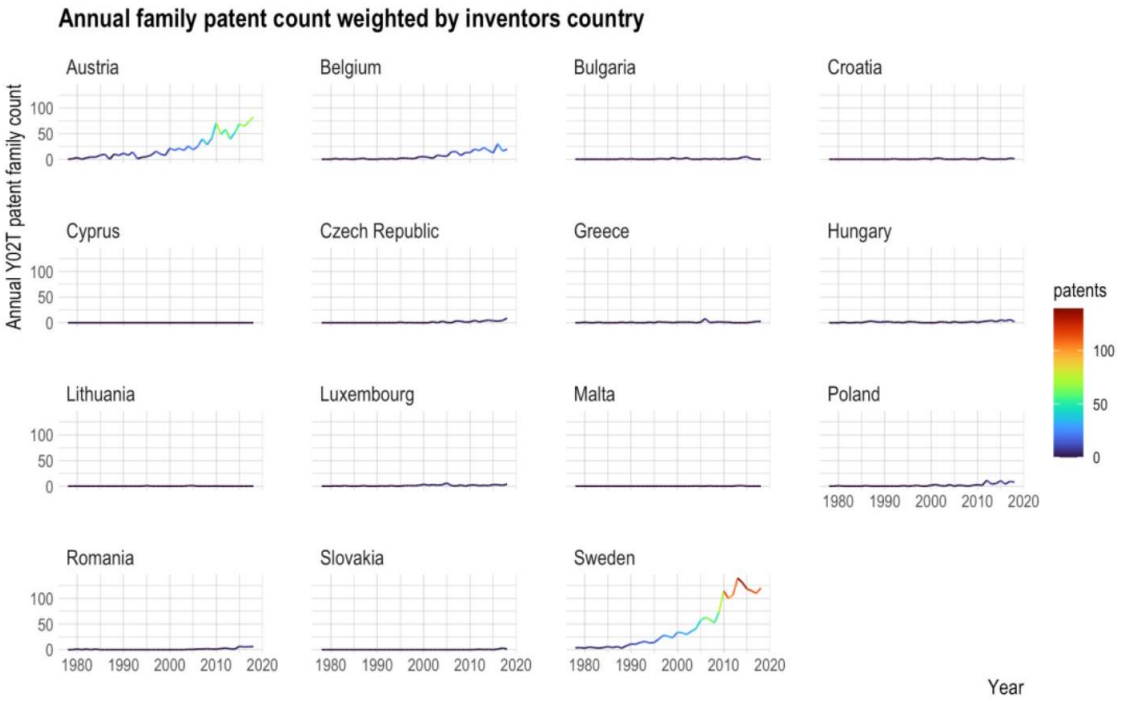


Figure 2: Trend Y02T patents by country from 1978 until 2018 (European Patent Office, 2022)

Table 1: Predictors of YO2T patents

Predictor	Source
Total climate change mitigation patents	OECD Statistics
Total triadic patents	OECD Statistics
Transport (WIPO) IP5 patent family	OECD Statistics
Y02T patents	PATSTAT
Climate transport EPO	OECD Statistics
B60 (WIPO) patents	OECD Statistics
GDP	Penn World Table
Enrollment in tertiary education	The World Bank

dicators to predict country-level innovation. GDP is stated as expenditure-side real GDP at chained PPP in millions of USD, with 2017 as the base year (Penn World Table, 2021). Enrollment in tertiary education is measured as the gross enrollment ratio in tertiary education of both sexes regardless of age (World Bank, 2016).

Y02T patents are those patents that concern climate mitigation technology in transportation. According to Abadie (2021), it is common to include the outcome variables as one of the predictors as this improves model results. This does not introduce a bias. The exact methodology behind Y02T patents is described in Chapter 3.1.

4. Results: A Large Gap between Sweden and Synthetic Sweden

In this section, I first present my findings concerning the analysis on the effect of the introduction of the carbon tax in 1990 on clean innovation in the transport sector in Sweden. Next, I conduct tests to verify that my results are not driven by chance.

4.1. Results of the Synthetic Control Method

Implementing a simple DiD regression for the data analyzed is unfeasible. In the case of a DiD regression, a bias is introduced as the parallel trends assumption is violated. Appendix 5 plots the development in Y02T patent families per year for Sweden and the mean of the donor pool countries. Panel (A) shows the years 1985 until 2018, while panel (B) zooms into the timeframe 1985 until 1995. Appendix 5 suggests that the equally weighted donor pool is unsuitable as a comparison group for Sweden to study the effect of the carbon tax on Y02T patents. In panel (B), the graph shows that the time series of Y02T patents in Sweden and the donor pool differs notably. There is no parallel trend. Hence, the implementation of a standard DiD model would be biased.

For the synthetic control method, no parallel trends assumption is needed. The central question to assess the effect of the introduction of the carbon tax in 1990 on Y02T patents in Sweden is how the Y02T patent trajectory would have developed after 1990 without the introduction of the carbon tax. As explained in chapter 3.2, I construct synthetic Sweden as the convex combination of countries in the donor pool, which are the closest resemblance of Sweden concerning the pre-policy values of Y02T predictors.

Table 2 shows the predictors included in the analysis and their respective weights assigned, which sum up to one. To test whether the result is robust when certain weights are not included in the analysis, I ran the entire synthetic control method multiple times, leaving the predictors out every time. The weights assigned to the countries constructing synthetic Sweden in the leave-one-out test remain unchanged. This indicates that predictors are similar as they predict the same countries to be chosen when creating synthetic Sweden. Including all of the predictors, triadic patent families and GDP receive the largest weights. The model computes that these predictors are best suited to minimize the mean squared prediction error (MSPE) of the synthetic control unit.

Table 3 provides a comparison between the mean values of Y02T predictors of Sweden, synthetic Sweden, and the donor pool mean. The values reported concern only the pre-treatment period from 1985 to 1989. Comparing the values between the three groups, the assumption that the mean of the donor pool itself is not well suited as a control group is corroborated. The mean values of the donor pool are significantly lower than those of Sweden. Although the values of synthetic Sweden are still lower than the ones of real Sweden, they are substantially higher and a much better approximation of actual Sweden during the pre-treatment period.

Appendix 6 displays the weights for each country in the donor pool. The outcome of the synthetic control model indicates that Y02T patents in Sweden pre-intervention can be most closely reproduced by a weighted combination of 63.8% Austria and 36.2% Belgium. All the other states are assigned a weight of 0%. As can be seen in Figure 2, the other countries in the donor pool, next to Austria and Belgium, have substantially fewer Y02T patents than Sweden. Similarly, the patent predictors during the pre-treatment period also have substantially lower values for the other countries next to Austria and Belgium. Hence, those countries are assigned zero percentage weights in the estimation of synthetic Sweden. According to the Bureau of Transport Statistics (2022), Sweden's annual number of vehicles produced is similar to that of Belgium and Austria, with 258, 224, and 125 thousand vehicles, respectively. Surprisingly, Slovakia and the Czech Republic have even larger vehicle exports, with over one million vehicles annually, but show fewer Y02T patent applications than Austria or Belgium. One possible explanation is that the innovation infrastructure in those countries is less established than in Austria or Belgium. Enrollment in tertiary education, one of my predictor variables for Y02T patents, is more than three times larger for Austria and Belgium than for Slovakia or the Czech Republic.

Appendix 7 shows the Y02T family patents trend from 1985 until 1995. During the pre-treatment period from 1985 until 1989, we can see that the trend of Y02T family patents of Sweden and synthetic Sweden is a better match than in Appendix 5, panel (B).

Figure 3 panel (A) plots the Y02T patent families for Sweden and its synthetic counterpart for the entire period analyzed. In panel (A), we can see that a large gap between Sweden and synthetic Sweden started to emerge around 1990. This explicit difference between the treated variable and the control group is plotted in panel (B). The estimate of the effect of the introduction of the carbon tax on innovation in clean transportation, Y02T patent families, in Sweden is the gap between Y02T patent families in Sweden and its synthetic counterpart after the introduction of the carbon tax. The strong positive trend of the gap in panel (B) suggests that the positive effect introduction of the carbon tax in 1990 on Y02T patents has a large magnitude. The results of the synthetic control method indicate that for the entire 1985-2018 period, Y02T patent family applications per year increased by 35.853 patents on average per year. The value of 35.853 is calculated by taking the mean of all the gap values from 1990 until 2018. Compared to the synthetic control group, I estimate that Sweden has 2.555 times more patents from 1990 until 2018. However, the estimated result overstates the actual effect of the carbon tax introduction on clean technology patents because in 1990, Sweden also broadened the coverage of its existing VAT, levied at 25%, to include transport fuels (Andersson, 2019). Andersson (2019) and Moore et al. (2021) disentangle the effect of the carbon tax and the VAT on the price elasticities of gasoline demand. Andersson (2019) uses a time-series analysis concerning the tax-exclusive price of gasoline, which refers to the gas price subtracted by the carbon tax on consumption. Moore et al. (2021) also use a time-series analysis and estimate the effect of the different fuel price components on clean patents. They both find that the effect of the carbon tax is larger than the effect of the VAT. More specifically, Moore et al. (2021) estimate that the effect of the carbon tax is double as large as that of the carbon-tax exclusive gasoline price. Therefore, the effect of the carbon tax is larger than that of the VAT on innovation in clean technology. Due to data availability concerning historical data on the gasoline and VAT price in Sweden, I cannot conduct such a time-series analysis. Nevertheless, Andersson's (2019) and Moore et al.'s (2021) findings on VAT will likely hold for my period analyzed.

My estimated result of the effect of the introduction of the carbon tax is significantly larger than Moore et al.'s (2021) analysis. Moore et al. (2021) estimate an average increase of 7.37 per year. However, they investigate the effect of introducing the carbon tax in Sweden on Y02T patents from 1990 until 1999. According to Figure 3, the patent frequency increased substantially after 1999, which certainly is one reason my estimation is larger.

Table 2: Y02T patent predictors and their weights

Predictor	Weight
Total triadic patents	0.361
GDP	0.321
Total climate change mitigation patents	0.186
Climate transport EPO	0.048
Enrollment in tertiary education	0.04
Y02T patents	0.036
Transport (WIPO) IP5 patent family	0.008
B60 (WIPO) patents	0.001

Table 3: Mean values of Y02T patent predictors

Predictors	Sweden	Synthetic Sweden	Donor Pool Mean
Transport (WIPO) IP5 patent family	57.705	26.702	4.153
Total triadic patents	422.515	188.038	31.896
Y02T patents	6.250	5.205	0.792
Total climate change mitigation patents	33.257	31.487	4.690
Climate transport EPO	6	5.283	0.830
Enrollment in tertiary education	30.368	30.717	17.494
GDP	239,209.000	202,904.200	134,954.100
B60 (WIPO) patents	27.519	14.498	2.365

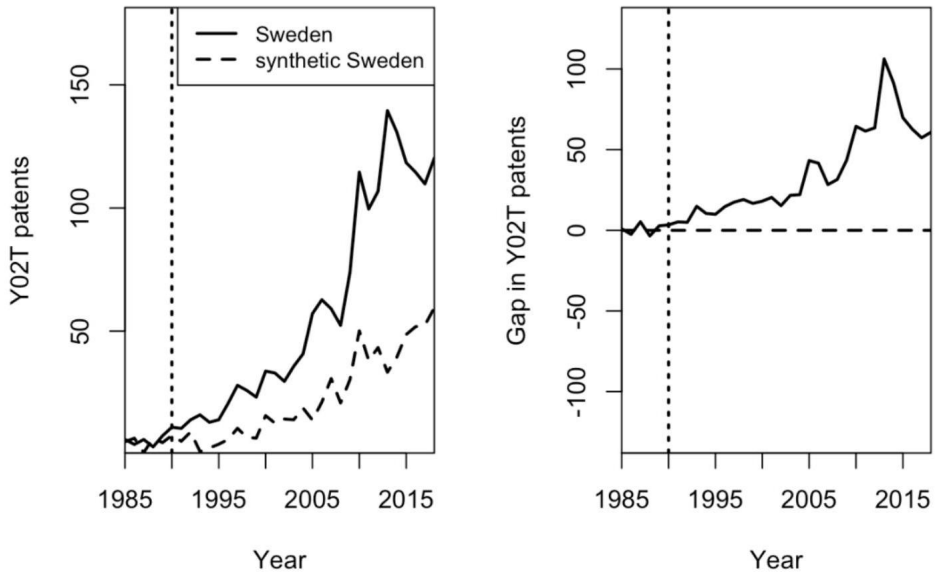


Figure 3: Trends and gap in Y02T patents: Sweden vs. synthetic Sweden

4.2. Inferences: The Result is Not Driven by Change

According to Abadie et al. (2010), the statistical significance of the estimates presented in the previous chapter is to be evaluated by answering whether the outcome might be driven by chance. I answer the question: How often would I obtain an outcome of at least the same magnitude if I had chosen another country in my donor pool instead of Sweden? I use the same methodology as Bertrand et al. (2004), Abadie and Gardeazabal (2003), and Abadie et al. (2010) by running placebo tests. I apply the same synthetic control algorithm to countries in the donor pool. Those countries did not implement a carbon tax during the period analyzed (1985-2018). If the placebo tests show that the gap estimated for Sweden is extraordinarily large, I interpret the the analysis provides significant evidence that the introduction of the carbon tax has a positive effect on Y02T patent applications in Sweden. However, if the placebo runs create similar gaps to the one

observed for Sweden, I cannot conclude that my analysis provides significant evidence.

According to Lehmann and Romano (2005), running placebo tests is analogous to conducting permutation tests in which the distribution of the test statistic is modeled by random permutations. Here the sample unit is assigned to the treatment and untreated groups.

I run placebo tests by applying the synthetic control method iteratively to the countries in the donor pool. This allows me to assess the significance of my estimates. In every interaction in the loop, I reassign the introduction of the carbon tax to one of the 14 countries in the donor pool and estimate the effect associated with the respective iteration. My result is a plot that shows me the trajectories of estimated gaps for the countries in which no carbon tax was introduced in reality.

Figure 4 shows the outcomes of the placebo test. The black line shows the estimated gap for Sweden, while the gray lines show the gap for each country in the donor pool and their created synthetic control group. It can be seen that the estimated gap for Sweden from 1990 until 2018 is large compared to the distribution of the gaps of the countries in the donor pool. Moreover, it is shown that the Y02T family patents for the period 1985 until 1990 can be well-reproduced for the other country by the convex combination. Placebo tests with a dissimilar fit prior to the policy intervention do not contribute to understanding the relative rarity of estimating a sizeable gap after the intervention for a country that actually had a good fit pre-intervention. One approach is to recreate Figure 4 by continually lowering the MSPE limit. By default, the MSPE limit is set to 20. Hence, only countries with a MSPE for the pre-treatment period of maximally 20 times larger than the treatment country are included. Abadie et al. (2010), for instance, lower the threshold until a value of 2 times the MSPE of the treatment country. In their case, some of the units investigated have very large MSPEs of over 1000, which create noise in the plot. In my analysis, I do not have countries with such a large MSPE. Therefore, only a slight visual difference can be seen when the MSPE limit is lowered. Lowering the MSPE limit to 5 results in the upper grey line disappearing in Figure 4. Further lowering the MSPE limit does not result in further changes.

Another popular method to evaluate the rarity of the gap analyzed for the treatment country compared to the countries in the donor pool is to look at the distribution of the pre-treatment MSPE ratio. A high ratio is desirable as it indicates a relatively small pre-treatment prediction error, indicating a good synthetic control, and a high post-treatment MSPE, indicating a large gap between the treated unit and its synthetic control after the treatment. However, as some countries in the donor pool do not have any Y02T patents during the pre-treatment period, the MSPE ratio test is biased. For example, Romania has not had Y02T patents during the pre-treatment period. Running the placebo for Romania, its synthetic control group is formed by the Czech Republic and Slovakia. Like Romania, The Czech Republic and Slovakia both had no Y02T patents during the pre-treatment period. The MSPE is very low for Romania during the pre-treatment period because synthetic Romania almost perfectly matches Romania. The distribution of the pre-treatment MSPE ratio is non-informative if countries with zero values during the pre-treatment period are present.

I bootstrap the synthetic control by drawing sub-samples from the donor pool. I conduct this 500 times to get the average placebo treatment effects distribution. I plot the distribution of the actual average treatment effect on the treated (ATT) with a dashed vertical line. The result of this bootstrapping is shown in Appendix 8. The distribution of the actual ATT is on the right and far away from most of the placebo ATTs, indicating that the outcome is substantially different from the placebo ATTs.

5. Discussion and Limitations

In this chapter, the results of my analysis are critically evaluated and scrutinized. Furthermore, I contextualize my results in the literature reviewed and present future research opportunities.

The results of my analysis on the effect of the introduction of the carbon tax in Sweden in 1990 on clean innovation have an economically relevant magnitude and are significant. My result is that there are 35.853 more clean transportation patents per year in Sweden than in a hypothetical scenario in which no carbon tax has been introduced. My findings align with those of Moore et al. (2021), who analyze the introduction of the carbon tax in Sweden on clean patents from 1990 until 1999. Similarly to my findings, they record a positive increase in clean patent frequency due to the introduction of the carbon tax. They find an average increase of 7.37 per year. As mentioned in chapter 4, my finding likely has a larger magnitude because the patent frequency of countries such as Austria, Belgium, and Sweden increased substantially after 1999. Moore et al. (2021) use a donor pool comprising 15 selected OECD countries, of which Belgium, France, Spain, and the United States receive a weight. Presumably, OECD countries were used for the donor pool because of high data availability. However, some OECD countries might be less than ideal to include in the donor pool when creating the counterfactual. OECD countries such as the United States or Japan are geographically distant, have a substantially larger population, and a history of filing the largest amount of patents at the main patent offices in the world (WIPO, 2022). The absolute number of patents issued by the United States or Japan is more than ten times larger than that of Sweden. While this might not have significantly impacted Moore et al. (2021) analysis, as it only reaches until 1999, it most certainly introduced a bias when conducting the analysis during a more recent period. If one chooses to include the United States or Japan in the analysis, one should find a suitable method to control for the significant innovative infrastructure that exists in those countries. Furthermore, Lee et al. (2011) find evidence that the United States auto industry had a strong innovation response to the United States performance-based technology-forcing

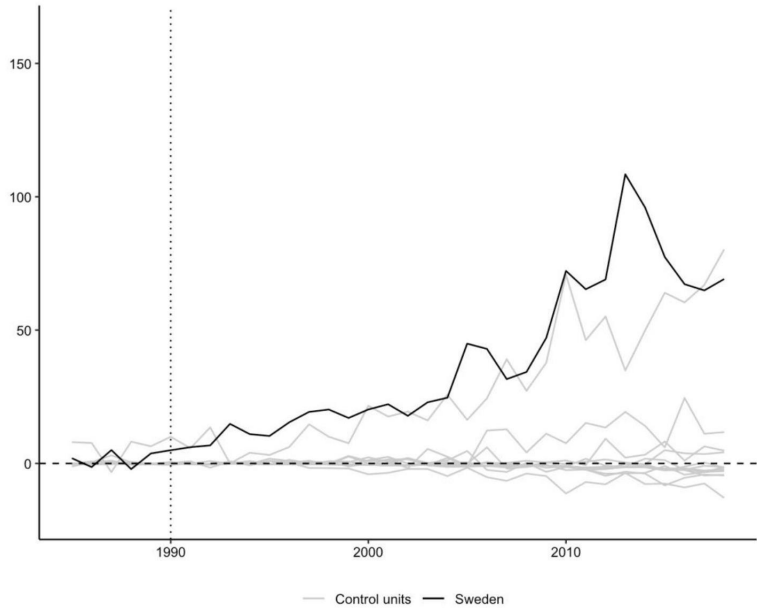


Figure 4: Difference between actual and synthetic countries (gaps) for Sweden (black) and control countries (grey)

command-and-control regulation between 1970 and 1998. This policy is not accounted for in Moore et al.'s (2021) analysis, although it influenced the patent frequency for the United States, one of the countries which constitutes a part of their synthetic Sweden.

Calel and Dechezleprêtre (2016), who investigate the effect of the EU ETS on low-carbon innovation, also find that the policy intervention increased low-carbon innovation among regulated companies compared to non-regulated ones.

As the preferred method of analysis, I choose the synthetic control method. The synthetic control method is especially well suited as I have only one treatment unit and aggregate data. In this case, the synthetic control method provides advantages over the DiD estimator. First, by implementing the synthetic control method, I can relax the parallel trends assumption of the DiD estimator by permitting the effects of potential confounding variables to change over time (Abadie et al., 2010). Second, I use predictors of Y02T patents to create a synthetic Sweden out of countries from the donor pool. The addition of the covariates total climate change mitigation-related patents, total triadic patents, transport IP5 patents, climate transport patents filed at the EPO, B60 patents, GDP, or ratio of people enrolled in tertiary education is not possible when using the DiD regression. These covariates are likely affected by the carbon tax implementation or are outcome variables themselves. Hence, they are considered suboptimal controls for the DiD regression. In the synthetic control method, I use the predictors to produce synthetic Sweden without creating a confounding effect during the post-treatment period (Andersson, 2019). Third, according to Card (1990) and Lin and Li (2011), the ambiguity when choosing comparison units is reduced as the synthetic control method chooses them through a data-driven method while they have to be chosen manually in the construction of the DiD method.

Although the relationship between the carbon tax price and patents is not the focus of my research, I analyze whether a larger carbon tax has a more significant impact on patents than a smaller one. To this end, I conduct a simple OLS regression with the annual Y02T patent count as my dependent variable and the carbon tax price as my independent variable. The regression is conducted during the period 1990 to 2018 and is specified as follows:

$$Y02T_t = \beta_1 * x + \varepsilon_t \tag{5}$$

Appendix 9 presents the regression results from estimating equation 5. The dependent variable, Y02Tt, depicts Sweden's absolute annual number of Y02T patents, and the independent variable represents the yearly carbon tax rate. The result is a slope on Y02T patents of .715 (Std. Error = .067), indicating that there are, on average, .715 Y02T patents more per year in Sweden for a one unit increase in the carbon tax, ceteris paribus. To assess whether the OLS assumptions are met, I assess whether nonlinearity is present with a residual plot in Appendix 10. Appendix 10 (left) shows that the relationship between the carbon tax price and the Y02T patents is linear. Appendix 10 (right) affirms this by plotting the residual with the geometric smooth fitting function using the loess method. Appendix 12 plots the residuals against the normal distribution and presents that the gaussianity assumption holds. The diagnostics plot in Appendix 12 shows that heteroscedasticity is not present. Hence, the i.i.d. assumption of the errors is not violated.

To assess whether my findings may apply to the transport sector in other countries, I carefully examine Sweden's political and social characteristics and determine how they might

differ from those of other countries.

Although Sweden is among the smaller-sized countries within the EU, it has often formed coalitions with other countries such as Austria, Belgium, Denmark, Finland, Luxembourg, and the Netherlands to support aggressive climate actions at the EU level (Schreurs & Tiberghien, 2007). Moreover, Sweden has been recognized as the country in the EU that has engaged in the largest number of climate policy initiatives (Burck et al., 2006). Sweden has been able to construct and implement environmental policies designed to be effective throughout the course of decades. 'New politics', a social and political movement concerning, inter alia, environmental pollution, strongly influenced Swedish politics in the 1970s and 1980s. In Sweden, this new social movement is characterized as being especially pragmatic and consensus-oriented, whereas in Germany student movements such as the Außerparlamentarische Opposition were opposed to the state and other larger institutions. While 'new politics' in most EU countries rarely got their voices heard politically, the movement got political attention in Sweden. It was even allowed to participate in the design of energy politics (Jahn, 1993).

Furthermore, Sweden's stable political system, which was represented by the Social Democratic Party (SAP), the Moderate Party (M), and the Left Party (V) since the 1920 s, created credibility, showed political commitment, and created a good government reputation (Jahn, 1993). Policies originating from a government that consistently acts in a credible manner have especially strong thrust as firms realize that policies are meant to stay. Thus, firms are able to make the necessary long-term strategic investments to respond accordingly (Brunner et al., 2012). However, it must be noted that political stability has decreased in recent years as the right-wing populist party Sweden Democrats (SD) managed to get a significant number of seats in the Swedish Riksdag (Jylhä et al., 2019).

According to G. Hofstede and Bond (1984), a country's culture can be described by a model assessing the characteristics: Power Distance, Individualism-Collectivism, Masculinity-Femininity, Uncertainty Avoidance, Long-Term Orientation, and Indulgence on a scale from 0-100. Shane (1995) shows that national differences concerning the ability to innovate do not only result from economic factors such as the industry structure, infrastructure, or societal welfare but also from cultural values people hold. Shane (1995) points out that uncertainty-accepting societies are more innovative than uncertainty-avoiding societies because roles such as transformational leaders have greater legitimacy. In Hofstede et al.'s (1984) model Sweden, Denmark, the Netherlands, and Finland score especially low on uncertainty avoidance (Hofstede, 2017). Simultaneously, these countries have especially high research systems, R&D expenditure, venture capital financing, public-private collaborations, and patent applications (Serafeim, 2015).

The social and political considerations show that Sweden indeed occupies a unique position as it scores low on uncertainty avoidance, green social movements influence politics, and the political system possesses sufficient stability to allow companies to respond adequately to climate policies. Therefore, introducing a carbon tax in other countries in the EU will likely not have as large of a positive effect on clean innovation. Nevertheless, as countries in the EU must follow the ESR, there is a new urgency to implement effective climate policy. This might help overcome obstacles such as high uncertainty avoidance. Moreover, environmental citizenship, which entails the right to participate in creating environmental policies, increasingly gains support through climate youth movements around Europe, such as Friday's for future (Wahlström et al., 2019).

The result found for the transport sector is to be compared to other sectors of the economy under careful consideration of the following factors. First, the transport sector in Sweden is of high materiality as it is an important constituent of the economy and a large employer (OEC, 2022). Other sectors which are of less economic importance might exhibit less public-private cooperation and, subsequently, less patent applications. Second, internationalized firms, which are present in the transport sector, file more patents than companies with less international presence. Internationalized firms see patents as a strategic protection against imitation (Neuhäusler, 2012). Third, infrastructure sectors like energy, sanitation, transportation, or water supply differ from other sectors in terms of significant degrees of capital intensity, sector-specific regulation, and long-lived assets. These characteristics create barriers to change and often result in path-dependent, incremental improvements rather than disruptive innovation. Hence, when applied to sectors that score low on the before mentioned characteristics, the carbon tax might have a more considerable impact on innovation than on the transport sector, which is considered to possess the aforementioned barriers to change (Markard, 2011). Summarizing, one might compare the result found to other sectors but most always contextualize the results and estimate the effect of innovation-inducing and inhibiting factors.

One limitation of my study is that the pre-treatment period only stretches from 1985 until 1989. Therefore, it cannot be verified over a prolonged period whether Sweden and synthetic Sweden follow the same trend pre-treatment. A more extended pre-treatment period might allow for even more precise construction of synthetic Sweden as the model can use a longer time series to create synthetic Sweden. Since I use OECD data for most of my predictor variables, which dates back to 1985 earliest, I cannot construct a larger pre-treatment period. The availability of a dataset dating back five to ten years might help construct an even closer matching synthetic Sweden.

One possible confounding factor might be that those companies impacted by the carbon tax might not innovate themselves but procure clean technology from a third party. This might bias my results because I measure innovation based on the investor's country of residence. While it is challenging to compute the extract effect this confounding variable might have, Fischer et al. (2003) and Milliman and Prince (1989)

estimate that developing proprietary technology is often of strategic advantage to companies compared to purchasing it from others.

Another caveat of my study is that I can only analyze patent data from PATSTAT until 2018. Patent applications are continuously filled to the large patent offices but only gradually uploaded to the PATSTAT database. Therefore, not every patent has been uploaded for the most recent years.

6. Conclusion and Opportunities for Future Research

This paper empirically shows that a carbon tax policy can successfully drive clean innovation. Clean innovation is important for long-term GHG emission reduction. Therefore, a carbon tax is an important tool to achieve the target under the Paris Climate Agreement to limit GHG emissions to well below 2°C compared to pre-industrial levels. In my empirical ex-post analysis, I find that after Sweden implemented a carbon tax and VAT on transport fuels, clean patents in the transportation sector increased by an average of 35.853 patent annually, compared to a scenario without a carbon tax, from 1990 until 2018. In my analysis, I do not disentangle the effect of the VAT due to data availability. However, Andersson's (2019) and Moore et al.'s (2021) studies indicate that the effect of the carbon tax on innovation is significantly stronger than that of the VAT. Although the exact magnitude of my finding is not directly generalizable to other countries or periods, my analysis shows that introducing a carbon tax can significantly affect clean innovation. As only a small number of predominantly Nordic countries have introduced a carbon tax at a meaningful rate, it is now at the time that other European governments as well consider implementing a carbon tax. This would allow them to achieve innovation and subsequently lower GHG emissions in the long run for those sectors that the EU ETS does not cover. My result is in line with Andersson's (2019) and Moore et al.'s (2021) findings, as they also find a positive and economically significant effect. I find a larger magnitude in my study, which is likely the case as I investigate a more recent period in which patent frequency has increased.

In my identification strategy, I carefully construct synthetic Sweden out of a donor pool of countries that did not implement a carbon tax or comparable policies. I show that my synthetic Sweden is a better counterfactual than the average of the donor pool countries during the pre-treatment period. Synthetic Sweden reproduces actual Sweden based on a set of predictors of Y02T patents. The results I obtain are robust, which I show through placebo tests and bootstrapping. Opportunities for future research are especially present concerning the moderating effect of clean innovation on GHG emissions. While there are numerous studies on the effect of carbon pricing on GHG emissions and the effect on innovation now has been estimated it is to be determined how strong the moderating effect of innovation on GHG emission is.

References

Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, *59*(2), 391–425.

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, *105*(490), 493–505. https://doi.org/10.1198/jasa.2009.ap08746

Abadie, A., Diamond, A., & Hainmueller, J. (2014). Comparative politics and the synthetic control method. *American Journal of Political Science*, *59*(2), 495–510. https://doi.org/10.1111/ajps.12116

Abadie, A., & Gardeazabal, J. (2003). The Economic Costs of Conflict: A Case Study of the Basque Country. *American Economic Review*, *93*(1), 113–132. https://doi.org/10.1257/000282803321455188

Acemoglu, D., Aghion, P., Bursztyn, L., & Hemous, D. (2012). The Environment and Directed Technical Change. *American Economic Review*, *102*(1), 131–166. https://doi.org/10.1257/aer.102.1.131

Aghion, P., Dechezleprêtre, A., Hemous, D., Martin, R., & Van Reenen, J. (2016). Carbon Taxes, Path Dependency, and Directed Technical Change: Evidence from the Auto Industry. *Journal of Political Economy*, *124*(1), 1–51.

Ambec, S., Cohen, M. A., Elgie, S., & Lanoie, P. (2013). The Porter Hypothesis at 20: Can Environmental Regulation Enhance Innovation and Competitiveness? *Review of Environmental Economics and Policy*, *7*(1), 2–22. https://doi.org/10.1093/reep/res016

Andersson, J. J. (2019). Carbon Taxes and CO2 Emissions: Sweden as a Case Study. *American Economic Journal: Economic Policy*, *11*(4), 1–30. https://doi.org/10.1257/pol.20170144

Angelucci, S., Hurtado-Albir, F. J., & Volpe, A. (2018). Supporting Global Initiatives on Climate Change: The EPO's "Y02-Y04S" Tagging Scheme. *World Patent Information*, *54*, S85–S92. https://doi.org/10.1016/j.wpi.2017.04.006

Arthur, W. B. (1989). Competing Technologies, Increasing Returns, and Lock-In by Historical Events. *The Economic Journal*, *99*(394), 116. https://doi.org/10.2307/2234208

Atradius. (2019). Market Monitor Automotive Sweden 2019. https://group.atradius.com/publications/market-monitor-automotive-sweden2019.html

Azam, A., Rafiq, M., Shafique, M., Zhang, H., & Yuan, J. (2021). Analyzing the Effect of Natural Gas, Nuclear Energy and Renewable Energy on GDP and Carbon Emissions: A Multi-Variate Panel Data Analysis. *Energy*, *219*, 119592. https://doi.org/10.1016/j.energy.2020.119592

Baumol, W. J., & Oates, W. E. (1988). *The Theory of Environmental Policy*. Cambridge University Press. https://doi.org/10.1017/cbo9781139173513

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How Much Should We Trust Differences-in-Differences Estimates? *The Quarterly Journal of Economics*, *119*(1), 249–275. https://doi.org/10.1162/003355304772839588

Bohlin, F. (1998). The Swedish Carbon Dioxide Tax: Effects on Biofuel Use and Carbon Dioxide Emissions. *Biomass and Bioenergy*, *15*(4-5), 283–291. https://doi.org/10.1016/s0961-9534(98)00036-1

Borghesi, S., & Montini, M. (2016). The Best (and Worst) of GHG Emission Trading Systems: Comparing the EU ETS with Its Followers. *Frontiers in Energy Research*, *4*. https://doi.org/10.3389/fenrg.2016.00027

Brunner, S., Flachsland, C., & Marschinski, R. (2012). Credible Commitment in Carbon Policy. *Climate Policy*, *12*(2), 255–271. https://doi.org/10.1080/14693062.2011.582327

Brunnermeier, S. B., & Cohen, M. A. (2003). Determinants of Environmental Innovation in US Manufacturing Industries. *Journal of Environmental Economics and Management*, *45*(2), 278–293. https://doi.org/10.1016/s0095-0696(02)00058-x

Burck, J., Bals, C., Treber, M., & Avram, R. (2006). *Climate Change Performance Index 2006*. https://www.germanwatch.org/sites/default/files/publication/1470.pdf

Bureau of Transportation Statistics. (2022). *World motor vehicle production, selected countries*. Retrieved October 16, 2022, from https://

www.bts.gov/content/worldmotor-vehicle-production-selected-countries

Calel, R., & Dechezleprêtre, A. (2016). Environmental Policy and Directed Technological Change: Evidence from the European Carbon Market. *Review of Economics and Statistics*, *98*(1), 173–191. https://doi.org/10.1162/rest_a_00470

Card, D. (1990). The impact of the mariel boatlift on the miami labor market. *Ilr Review*, *43*(2), 245–257.

Chen, Y.-h., Wang, C., Nie, P.-y., & Chen, Z.-r. (2020). A Clean Innovation Comparison Between Carbon Tax and Cap-and-Trade System. *Energy Strategy Reviews*, *29*, 100483. https://doi.org/10.1016/j.esr.2020.100483

Cheng, Y., Sinha, A., Ghosh, V., Sengupta, T., & Luo, H. (2021). Carbon Tax and Energy Innovation at Crossroads of Carbon Neutrality: Designing a Sustainable Decarbonization Policy. *Journal of Environmental Management*, *294*, 112957. https://doi.org/10.1016/j.jenvman.2021.112957

Collier, U., & Löfstedt, R. E. (1997). Think globally, act locally?: Local climate change and energy policies in sweden and the uk. *Global environmental change*, *7*(1), 25–40.

Commission, H.-L. (2017). Report of the High-Level Commission on Carbon Prices.

Crabb, J. M., & Johnson, D. K. N. (2010). Fueling innovation: The impact of oil prices and cafe standards on energy-efficient automotive technology. *The Energy Journal*, 199–216.

Cunningham, S. (2021). *Causal Inference*. Yale University Press. https://doi.org/10.12987/9780300255881

Dechezleprêtre, A., & Sato, M. (2017). The Impacts of Environmental Regulations on Competitiveness. *Review of Environmental Economics and Policy*, *11*(2), 183–206. https://doi.org/10.1093/reep/rex013

Elgie, S., & McClay, J. (2013). Policy Commentary/Commentaire BC's Carbon Tax Shift Is Working Well After Four Years (Attention Ottawa). *Canadian Public Policy*, *39*(Supplement 2), S1–S10. https://doi.org/10.3138/cpp.39.supplement2.s1

European Commission. (2022a). *Effort Sharing 2021-2030: Targets and Flexibilities*. Climate Action. https://climate.ec.europa.eu/eu-action/effort-sharing-member-states-emission-targets/effort-sharing-2021-2030-targets-and-flexibilities_en

European Commission. (2022b). *European Innovation Scoreboard*. https://ec.europa.eu/assets/rtd/eis/2022/ec_rtd_eis-country-profile-se.pdf

European Environment Agency. (2021). *EU-27 GHG Emission Trends and Projections Under the Scope of the Effort Sharing Legislation*. https://www.eea.europa.eu/data-and-maps/figures/eu-27-ghg-emissiontrends

European Patent Office. (2022). *PATSTAT Global 2022 Edition*.

Fischer, C., Parry, I. W. H., & Pizer, W. A. (2003). Instrument choice for environmental protection when technological innovation is endogenous. *Journal of Environmental Economics and Management*, *45*(3), 523–545.

Frietsch, R., & Schmoch, U. (2009). Transnational patents and international markets. *Scientometrics*, *82*(1), 185–200. https://doi.org/10.1007/s11192-009-0082-2

Fuss, S., Lamb, W. F., Callaghan, M. W., Hilaire, J., Creutzig, F., Amann, T., Beringer, T., de Oliveira Garcia, W., Hartmann, J., Khanna, T., et al. (2018). Negative emissions—part 2: Costs, potentials and side effects. *Environmental research letters*, *13*(6), 063002.

Gerlagh, R. (2008). A climate-change policy induced shift from innovations in carbon-energy production to carbon-energy savings. *Energy Economics*, *30*(2), 425–448. https://doi.org/10.1016/j.eneco.2006.05.024

Goulder, L. H., & Schein, A. R. (2013). Carbon taxes versus cap and trade: a critical review. *Climate Change Economics*, *04*(03), 1350010. https://doi.org/10.1142/s2010007813500103

Goulder, L. H., & Schneider, S. H. (1999). Induced technological change and the attractiveness of CO2 abatement policies. *Resource and Energy Economics*, *21*(3-4), 211–253. https://doi.org/10.1016/s0928-7655(99)00004-4

Hagedoorn, J., & Cloodt, M. (2003). Measuring innovative performance: Is there an advantage in using multiple indicators? *Research policy*, *32*(8), 1365–1379.

Hammar, H., & Åkerfeldt, S. (2011). CO2 Taxation in Sweden - 20 Years of Experience and Looking Ahead.

Harhoff, D., Narin, F., Scherer, F. M., & Vopel, K. (1999). Citation frequency and the value of patented inventions. *Review of Economics and statistics*, *81*(3), 511–515.

Harhoff, D., Scherer, F. M., & Vopel, K. (2003). Citations, family size, opposition and the value of patent rights. *Research Policy*, *32*(8), 1343–1363. https://doi.org/10.1016/s0048-7333(02)00124-5

Hascic, I., de Vries, F., Johnstone, N., & Medhi, N. (2009). Effects of Environmental Policy on the Type of Innovation: The Case of Automotive Emission Control Technologies. *Economic Studies*.

Hicks, J. (1932). *The theory of wages*. Springer.

Hofstede. (2017). *Home - Hofstede Insights organisational culture consulting*. https://www.hofstede-insights.com

Hofstede, G., & Bond, M. H. (1984). Hofstede's culture dimensions. *Journal of Cross-Cultural Psychology*, *15*(4), 417–433. https://doi.org/10.1177/0022002184015004003

Howard, P. H., & Sylvan, D. (2015). The economic climate: Establishing expert consensus on the economics of climate change. *Institute for Policy Integrity*, 438–441.

IISD. (2021). *Trading begins under China's National ETS*. https://sdg.iisd.org/news/trading-begins-under-chinas-national-ets/

International Energy Agency. (2007). *Climate Policy Uncertainty and Investment Risk*.

International Energy Agency. (2022). *Sweden carbon tax*. IEA/IRENA Renewables Policies Database. https://www.iea.org/policies/12725-sweden-carbon-tax

Jaffe, A. B., Newell, R. G., & Stavins, R. N. (2003). Technological change and the Environment. In *Environmental Degradation and Institutional Responses* (pp. 461–516). Elsevier. http://dx.doi.org/10.1016/s1574-0099(03)01016-7

Jahn, D. (1993). The rise and decline of new politics and the Greens in Sweden and Germany. *European Journal of Political Research*, *24*(2), 177–194. https://doi.org/10.1111/j.1475-6765.1993.tb00375.x

Johansson, B. (2000). Economic instruments in practice 1: Carbon tax in Sweden. *Workshop on Innovation and the Environment*.

Jonsson, S., Ydstedt, A., & Asen, E. (2020). Looking Back on 30 Years of Carbon Taxes in Sweden. *The Tax Foundation*. https://files.taxfoundation.org/20200929121706/Looking-Back-on-30-Years-of-Carbon-Taxes-in-Sweden-re.pdf

Jylhä, K. M., Rydgren, J., & Strimling, P. (2019). Radical right-wing voters from right and left: Comparing Sweden Democrat voters who previously voted for the Conservative Party or the Social Democratic Party. *Scandinavian Political Studies*, *42*(3-4), 220–244. https://doi.org/10.1111/1467-9477.12147

Karakaya, E., Nuur, C., & Assbring, L. (2018). Potential transitions in the iron and steel industry in Sweden: Towards a hydrogen-based future? *Journal of Cleaner Production*, *195*, 651–663. https://doi.org/10.1016/j.jclepro.2018.05.142

Khan, Z., Ali, S., Dong, K., & Li, R. Y. M. (2021). How does fiscal decentralization affect CO2 emissions? The roles of institutions and human capital. *Energy Economics*, *94*, 105060. https://doi.org/10.1016/j.eneco.2020.105060

Knittel, C. R., & Sandler, R. (2018). The welfare impact of second-best uniform pigouvian taxation: Evidence from transportation. *American Economic Journal: Economic Policy*, *10*(4), 211–242. https://doi.org/10.1257/pol.20160508

Lee, J., Veloso, F. M., & Hounshell, D. A. (2011). Linking induced technological change, and environmental regulation: Evidence from patenting in the U.S. auto industry. *Research Policy*, *40*(9), 1240–1252. https://doi.org/10.1016/j.respol.2011.06.006

Lehmann, E. L., & Romano, J. P. (2005). *Testing statistical hypotheses*. Springer.

Life Unify. (2022). Implementing the effort sharing regulation at national level - Life unify new report. https://caneurope.org/implementing-effort-sharing-regulation-at-national-levellife-unify-new-report/

Lilliestam, J., Patt, A., & Bersalli, G. (2020). The effect of carbon pricing on technological change for full energy decarbonization: A review of empirical ex-post evidence. *WIREs Climate Change*, *12*(1), 681. https://doi.org/10.1002/wcc

Lin, B., & Li, X. (2011). The effect of carbon tax on per capita CO2 emissions. *Energy Policy*, *39*(9), 5137–5146. https://doi.org/10.1016/j.enpol.2011.05.050

Lundberg, J. (2014). *The Right Price for Carbon Dioxide*. https://timbro.se/app/uploads/2017/01/ratt_pris_pa_koldioxid_101.pdf

Markard, J. (2011). Transformation of infrastructures: Sector characteristics and implications for fundamental change. *Journal of Infrastructure Systems*, *17*(3), 107–117. https://doi.org/10.1061/(asce)is.1943-555x.0000056

Metcalf, G. E., & Weisbach, D. A. (2009). The design of a carbon tax. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1324854

Milliman, S. R., & Prince, R. (1989). Firm incentives to promote technological change in pollution control. *Journal of Environmental Economics and Management*, *17*(3), 247–265.

Ministry of Finance. (2021). *Carbon Taxation in Sweden. Government Offices of Sweden*.

Moore, N., Brehm, J., & Gruhl, H. (2021). Driving Innovation? Carbon Tax Effects in the Swedish Transport Sector. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3995632

Murray, B., & Rivers, N. (2015). British columbia's revenue-neutral carbon tax: A review of the latest'grand experiment'in environmental policy.

Natur Vårds Verket. (2019). *Climate Statistics*. https://www.naturvardsverket.se/dataoch-statistik/klimat

Natur Vårds Verket. (2022a). *Fuel use for petrol and diesel cars*. https://www.naturvardsverket.se/data-och-statistik/trafik-ochtransporter/bransleanvandning-bensin-dieselbilar/

Natur Vårds Verket. (2022b). *Territorial and removals of greenhouse gases*. https://www.naturvardsverket.se/data-och-statistik/klimat/vaxthusgaserterritoriella-utslapp-och-upptag/

Neuhäusler, P. (2012). The use of patents and informal appropriation mechanisms: Differences between sectors and among companies. *Technovation*, *32*(12), 681–693. https://doi.org/10.1016/j.technovation.2012.07.004

Newell, R. G., Jaffe, A. B., & Stavins, R. N. (1999). The induced innovation hypothesis and energy-saving technological change. *The Quarterly Journal of Economics*, *114*(3), 941–975.

OEC. (2022). Sweden (SWE) exports, imports, and trade partners. *OEC - The Observatory of Economic Complexity*. https://oec.world/en/profile/country/swe

OECD. (2009). OECD Patent Statistics Manual. http://dx.doi.org/10.1787/9789264056442-en

OECD. (2016). *Share of emissions priced and average price signals from tax & ETS*. http://oe.cd/emissionsdata

OECD. (2018a). *Patents by Technology. OECD Statistics*. https://stats.oecd.org/Index.aspx?DataSetCode=PATS_IPC

OECD. (2018b). *SUPPLEMENT TO EFFECTIVE CARBON RATES*.

OECD. (2019). *Taxing Energy Use 2019. OECD*.

OECD. (2020). *OECD Science, Technology and Innovation Scoreboard*. https://www.oecd.org/sti/scoreboard.htm?i=G_XGDP%5C&v=6%5C&t=2010,2020%5C&s=OECD,SWE%5C&r=ci

OECD. (2022). *GDP per capita, USD, constant prices and PPPs, reference year 2005. OECD Statistics*. https://stats.oecd.org/

OECD Environment Directorate & International Energy Agency. (2001). GOOD PRACTICE GREENHOUSE ABATEMENT POLICIES: TRANSPORT. https://www.oecd.org/env/cc/2345641.pdf

Palmer, K., Oates, W. E., & Portney, P. R. (1995). Tightening environmental standards: The benefit-cost or the no-cost paradigm? *Journal of Economic Perspectives*, *9*(4), 119–132. https://doi.org/10.1257/jep.9.4.119

Patt, A., & Lilliestam, J. (2018). The case against carbon prices. *Joule*, *2*(12), 2494–2498. https://doi.org/10.1016/j.joule.2018.11.018

Pavitt, K. (1988). Uses and abuses of patent statistics. In *Handbook of quantitative studies of science and technology* (pp. 509–536). Elsevier.

Penn World Table. (2021). *Expenditure-side real GDP at chained. Federal Reserve Bank of St. Louis*. https://fred.stlouisfed.org/release?rid=285

Pigou, A. C. (1920). *The Economics of Welfare*. Macmillan.

Popp, D. (2002). Induced innovation and energy prices. *American Economic Review*, *92*(1), 160–180. https://doi.org/10.1257/000282802760015658

Porter, M. E. (1991). Essay. *Scientific American*, *264*(4), 168–168. https://doi.org/10.1038/scientificamerican0491-168

Raghupathi, V., & Raghupathi, W. (2017). Innovation at country-level: Association between economic development and patents. *Journal of Innovation and Entrepreneurship*, *6*(1). https://doi.org/10.1186/s13731-017-0065-0

Regeringskansliet. (2014). *Proposition*. Regeringskansliet. https://www.regeringen.se/rattsliga-dokument/proposition/1997/09

Rivers, N., & Schaufele, B. (2015). Salience of carbon taxes in the gasoline market. *Journal of Environmental Economics and Management*, *74*, 23–36. https://doi.org/10.1016/j.jeem.2015.07.002

SCB. (2020). *Sweden's research and development increased*. Statistiska Centralbyrån. https://www.scb.se/en/finding-statistics/statistics-by-subject-area/educationand-research/research/research-and-development-in-sweden/pong/statisticalnews/research-and-development-in-sweden-2019/

Schreurs, M. A., & Tiberghien, Y. (2007). Multi-Level reinforcement: Explaining European Union leadership in climate change mitigation. *Global Environmental Politics*, *7*(4), 19–46. https://doi.org/10.1162/glep.2007.7.4.19

Serafeim, G. (2015). Greece: Science and innovation profile of Greece. *Harvard Business Review*.

Shane, S. (1995). Uncertainty avoidance and the preference for innovation championing roles. *Journal of International Business Studies*, *26*(1), 47–68. https://doi.org/10.2307/155477

Shmelev, S. E., & Speck, S. U. (2018). Green fiscal reform in Sweden: Econometric assessment of the carbon and energy taxation scheme. *Renewable and Sustainable Energy Reviews*, *90*, 969–981. https://doi.org/10.1016/j.rser.2018.03.032

Stavins, R. N. (2007). A U.S. cap-and-trade system to address global climate change. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1026353

Teixidó, J., Verde, S. F., & Nicolli, F. (2019). The impact of the EU Emissions Trading System on low-carbon technological change: The empirical evidence. *Ecological Economics*, *164*, 106347. https://doi.org/10.1016/j.ecolecon.2019.06.002

Tybout, R. A. (1972). Pricing pollution and other negative externalities. *The Bell Journal of Economics and Management Science*, *3*(1), 252. https://doi.org/10.2307/3003077

United Nations. (2021). *Global Issues | Climate Change*. https://www.un.org/en/global-issues/climate-change

Unruh, G. C. (2000). Understanding carbon lock-in. *Energy Policy*, *28*(12), 817–830. https://doi.org/10.1016/s0301-4215(00)00070-7

Venmans, F. (2012). A literature-based multi-criteria evaluation of the EU ETS. *Renewable and Sustainable Energy Reviews*, *16*(8), 5493–5510. https://doi.org/10.1016/j.rser.2012.05.036

VINNOVA SE. (2022). *Vehicle strategic research and innovation*. Vinnova. https://www.vinnova.se/en/m/strategic-vehicle-research-and-innovation/aboutffi/

Vogt-Schilb, A., Meunier, G., & Hallegatte, S. (2018). When starting with the most expensive option makes sense: Optimal timing, cost and sectoral allocation of abatement investment. *Journal of Environmental Economics and Management*, *88*, 210–233. https://doi.org/10.1016/j.jeem.2017.12.001

Wahlström, M., Kocyba, P., De Vydt, M., & de Moor, J. (2019). Protest for a future: Composition, mobilization and motives of the participants in Fridays For Future climate protests on 15 March, 2019 in 13 European cities.

Weitzman, M. L. (1974). Prices vs. quantities. *The Review of Economic Studies*, *41*(4), 477. https://doi.org/10.2307/2296698

WIPO. (2022). *Interactive charts: IP facts and figures*. https://www.wipo.int/edocs/infogdocs/en/ipfactsandfigures/

World Bank. (2016). *The World Bank Education statistics*. https://datatopics.worldbank.org/education/

World Bank. (2022). *Pricing carbon*. https://www.worldbank.org/en/programs/pricing-carbon

Junior Management Science 9(1) (2024) 1306-1340

Junior Management Science

www.jums.academy
ISSN: 2942-1861

How to Measure the Success of Technology-Based Start-Ups – A Comprehensive Overview of the Perspectives of Academics & Practitioners

Faris Ben Saad

Technical University of Munich

Abstract

Successful technology start-ups can be a significant driver of a country's economic development and could transform entire industries with new technological innovations. For this reason, in research and practice, special attention is always paid to one particular type of start-up: a successful one. To date, however, little research has been done on how to measure a start-up's success. To advance the knowledge about start-up success measurement in academic research, this thesis investigates what academics and practitioners understand by a successful start-up and what they consider to be reliable measures of success. Several scientific studies dedicated to the examination of start-up success were analyzed and seven semi-structured expert interviews with venture capitalists from the early-stage investment sector were conducted. The results show that in both the academic and practical world, start-up success is perceived as a complex, multidimensional phenomenon whose measurement depends on a variety of different factors that may change over time. It is therefore concluded that a meaningful measurement of start-up success requires the use of a combination of different metrics to address this multidimensional nature of success.

Keywords: new venture performance; new ventures; startup success; startups; venture capital

1. Introduction

Technology start-ups can be a significant driver of a country's economic development. They create a large number of new jobs and can transform entire industries with new technological innovations (Christensen & Bower, 1996; M. Song et al., 2008). For this reason, start-ups are not only receiving growing attention in the practical but also in the academic world. Scientific interest in start-ups has increased substantially in the last two decades, with scholars from a diverse range of fields devoting their research to this topic (see Appendix A1).

Despite their considerable relevance for the economy and society, only a small fraction of all start-ups manages to survive in the long term (Schlichte et al., 2019). Technology start-ups in particular have comparatively low survival rates (M. Song et al., 2008). A considerable number of researchers is therefore concerned with finding out how to improve the survival chances of start-ups and increase their likelihood of success, resulting in a variety of different empirical studies (Jin et al., 2017; M. Song et al., 2008). An essential component of these studies is the measurement of the central dependent variable: success (Witt, 2004). To determine whether a start-up is successful, researchers have to measure success by means of selected metrics. The choice of the right success measures should be well considered, as they can have a decisive influence on the results and validity of a study (Eveleens et al., 2017).

First and foremost, I would like to extend my heartfelt thanks to my supervisor Riccarda Joas, who not only gave me the opportunity to write this thesis but also provided me with invaluable suggestions, constructive feedback, and brilliant ideas that greatly enhanced the quality of my work. I am truly grateful for her continuous support and mentorship throughout the entire writing process. Additionally, I would like to extend my appreciation to the venture capitalists who graciously shared their valuable time and knowledge with me during the qualitative research phase. Lastly, I want to extend my gratitude to Lisa Marie Hamacher for her unwavering support, encouragement, and belief in my abilities. Her continuous motivation and willingness to listen to my ideas have been instrumental in keeping me focused and motivated throughout this journey.

DOI: https://doi.org/10.5282/jums/v9i1pp1306-1340

The correct measurement of success is thus of major importance for researchers in the field of start-up success. Nevertheless, little research has been done on how to measure the success of start-ups. Given the ever-growing research interest in start-ups and the need for sound success measurement decisions, there is a great demand for further research on this topic (Eveleens et al., 2017; Kiviluoto, 2013). This academic work addresses this need, focusing specifically on technology-based start-ups. The aim of this thesis is to find out what academics as well as relevant practitioners define as a successful start-up and what they consider to be reliable measures of success. This is intended to contribute to a better understanding of the success phenomenon in start-up research and a more profound measurement of start-up success (Kiviluoto, 2013). Scholars, but also practitioners from the start-up environment, are to be shown with this thesis how the success of start-ups can be measured and what has to be considered when measuring success.

To achieve the objective of this academic work, a systematic literature review is conducted in the first part of this thesis, aiming to investigate the views of academics on the topic of success measurement. In order to capture the practitioners' perspective, this literature review is followed by an empirical study in which seven early-stage venture capitalists - who were considered relevant practitioners of a start-up ecosystem - are interviewed about the central research question.

2. Literature Review – Academics' Perspective

To capture the academics' perspective on the central research question, existing literature dedicated to the topics of start-ups and start-up success was analyzed. In the first half of this chapter, the review methodology used is described, and an extensive explanation of the start-up term and relevant concepts for this thesis is provided. This is followed by the main part of this chapter, which provides an overview of how start-up success is defined and measured in the scientific literature.

2.1. Review Methodology

In order to ensure an unbiased and comprehensive outcome, a systematic literature review was conducted (Tranfield et al., 2003). Following the approach of Crossan and Apaydin (2010), the review consisted of four distinct phases: planning, data collection, data analysis, and reporting. The first three phases are described below.

2.1.1. Planning

During the planning stage, the aim, as well as the scope of the literature review, was defined. The primary objective of the literature review was to provide a comprehensive overview of how start-up success is measured in entrepreneurship and management research and why certain metrics or measurement approaches are used. The literature review was limited to peer-reviewed scientific journals, as these are considered to be of high quality and influence for a scientific field (Crossan & Apaydin, 2010; Podsakoff et al., 2005). Due to their volume of peer-reviewed literature, the academic databases Scopus and ScienceDirect were used for the literature search. In addition, Google Scholar was employed to retrieve identified articles in PDF format.

2.1.2. Data collection and analysis

Prior to the collection of appropriate literature, the concept of start-up was examined in depth and finally defined due to its relevance to the topic of success measurement. To identify relevant studies, two separate searches were conducted: an initial broader search to obtain a general overview of existing success literature and a more specific search that was specifically designed to answer the central research question. To ensure a high literature quality, only studies published in influential entrepreneurship and management journals were considered in the search, i.e. the journal in which a study was published had to either be at least a C, preferably a B to A+ journal according to the VHB ranking or, alternatively have an impact factor of at least four.

In the initial search, the selected electronic databases were searched for scientific articles that were published within the last 40 years (1982-2022) and whose title, abstract, or keywords contained combinations of the terms "start-up", "startup", "venture", "success", "performance" and "measure*". Figure 1 shows the main query used in this process.

The initial search resulted in about 742 articles. After the exclusion criteria listed in Table 1 were applied to this sample by manually reviewing the individual title, abstract and/or conclusion of each article, the number of studies was reduced to 36.

To obtain a concrete overview of how academics measure the success of start-ups, the initial research was followed by a search for empirical studies that examine the success of start-ups and were published in the past 20 years (2000-2022). For this purpose, the query illustrated in Figure 2 was used.

During the search, a special focus was placed on young independent technology start-ups. Therefore, articles with a focus on corporate, public, or late-stage start-ups were intentionally excluded. Subsequently, all studies that met any of the criteria in Table 2 were also excluded.

As a result, 35 matching empirical studies were identified. Finally, a forward, as well as backward search, was again conducted based on the 71 identified studies to supplement the previously collected literature. The overall literature search resulted in a sample of 114 scientific articles, which were analyzed in the next step.

The analysis of the collected literature was based on the thematic codes listed in Table 3.

While 16 thematic codes were applied to the total of 42 empirical studies, only 7 to 8 more general codes were used for the remaining research articles.

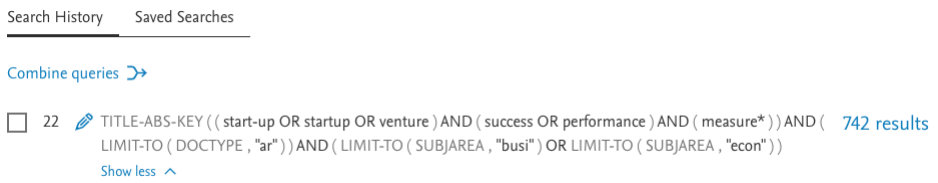


Figure 1: Scopus Query: Initial Search

Table 1: *Exclusion criteria: Initial Search*

Criteria
• Articles that were of empirical nature and provided no further relevant information
• Articles that focused only on late-stage start-ups
• Articles that did not conceptualize the concept of start-up success / performance
• Articles that were only concerned with entrepreneurial success

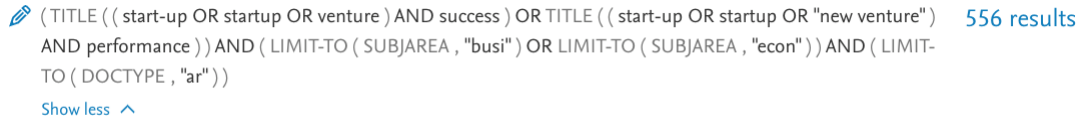


Figure 2: Scopus Query: Research Articles studying start-up success or performance

Table 2: *Exclusion criteria: Search for Success investigating Articles*

Criteria
• Articles specifying the type of success or performance in their title
• Articles that were not of empirical nature
• Articles on entrepreneurial success
• Articles that did not state properly how they measured success

Note. Furthermore, articles examining technology start-ups were preferred.

2.2. Start-up at a glance

In the further course of this thesis, it will be important to understand what a start-up is and which development phases such a company goes through. This chapter examines the concept of start-up by drawing on well-known start-up literature as well as high-impact entrepreneurship and management research.

2.2.1. Definitions of the start-up concept

The term "startup" was first mentioned in 1976 within a Forbes article published at the time to describe a newly formed business (Nguyen-Duc et al., 2020). The aspect of newness can also be found in more recent definitions, but it has become considerably less relevant, which can be seen in Table 4.

A review of existing definitions from selected entrepreneurship and start-up literature reveals that start-ups are increasingly ascribed characteristics that go beyond the purely time-related definition. It also became evident that there is no uniform characterization for a start-up, but that there are overlaps between existing definitions. Start-ups are particularly described as innovative, scalable, and growth-oriented (Tech, 2018). For example, Blank and Dorf (2012) state that a start-up is not an ordinary company but an organization looking for a suitable business model that will allow for growth and profitability. According to Graham (2012), especially the ability to grow disproportionately fast is the key criterion that differentiates a start-up from an ordinary company. In this context, he further adds that a start-up can only achieve this rapid growth if it is able to successfully serve a very large market (Graham, 2012). It thus becomes apparent that start-ups are not newly founded companies, but young organizations that aim for above-average growth in various business dimensions and therefore want to serve a large market (Tech, 2018). Research has shown that such a venture can take an average of eight to ten years to reach profitability. This time frame is therefore, often set as a threshold to characterize start-ups in terms of time (Davila et al., 2003; Li, 2020).

Table 3: *Thematic Codes for Literature Analysis*

Scope	Code
All Articles	Author name (year) Year Title Thesis Section Notes JQ3
23 Success Definition Articles	Success Definition
42 Empirical Studies	Sample Size Sample Type Company age (in years) #Employees Success Metrics Success Dimenison Number of Dimensions, [options: GROWTH, FUNDING, PROFITABILITY, LIQUIDITY EVENT, OPERATIONAL, EFFICIENCY, SURVIVAL, OTHER] Number of Measures Measure Type [options: f-financial, nf-non-financial, obj-objective, subj-subjective] Journal
Remaining Articles	Relevance Key point

2.2.2. Start-up in academic research

To obtain an overview of how start-ups are characterized in the research literature relevant to this thesis, a sample of 18 research articles from 12 different high-impact journals was reviewed (see Appendix A2). The corresponding articles were identified using the Scopus database by employing the search query illustrated in Figure 3.

Each article was dedicated to the study of start-ups and was therefore examined for both direct and indirect start-up characterizations. An analysis of the selected sample yielded 12 characteristics that were primarily used to describe start-ups. Table 5 illustrates these.

Four of these characteristics were applied particularly frequently: young, new, lack of resources, and high uncertainty. Thus, a large proportion of the sample defined a start-up mainly as a newly established firm with a relatively short operating history, which lacks resources and faces a high degree of uncertainty. Relatively little importance was attached to the characteristics of growth orientation and innovativeness. In fact, start-ups were explicitly distinguished from high-growth and innovative start-ups, with the latter two types mostly corresponding to the characterizations made in the previous chapter (Audretsch et al., 2021; Cacciolatti et al., 2020; S. Lee, 2022). This indicates that start-up definitions in research and practice slightly differ from each other. Within the reviewed academic literature start-up characterizations still seem to be predominantly influenced by the original meaning of the term. That is, a start-up is primarily considered a newly formed company and provided it has strong growth intentions and an innovative approach, it is specifically referred to as a high-growth or innovative start-up.

2.2.3. Defining a start-up

Due to the differences between various start-up definitions, it is necessary to define a start-up in concrete terms for the purposes of this thesis. Since both the academic and practical perspectives are of great relevance to this study, an attempt was made to incorporate both views in this definition. The term "start-up" is used in the further course of this work to describe a young company that (1) is not older than 10 years, (2) has strong growth intentions, (3) is developing innovative technology-based products or services and (4) strives for a scalable business model that enables profitability in the medium to long term.

2.2.4. Development stages of start-ups

In the course of their lifespan, start-ups develop on various interrelated levels and thus pass through different stages of development (Kumbhat & Sushil, 2018). A handful of researchers and practitioners have devoted themselves to identifying and explaining these developmental stages, resulting in a wide variety of models that have been created and published over the past several decades (Blank, 2007; Kazanjian, 1988; Kazanjian & Drazin, 1990; Kumbhat & Sushil, 2018; Marmer et al., 2011; Tech, 2018). Based on these models, a four-phase model was developed that explains the development of a start-up while sufficiently incorporating the financing rounds of the VC sector (see Table 6). The inclusion of the

Table 4: *Definitions of a Start-up*

Reference	Year	Title	Definition	Characteristics
Santisteban et al. (2021)	2021	Critical success factors for technology-based startups	[...] a small, dynamic, flexible, high-risk company that has a reproducible and scalable business model and provides innovative products and/or services.	small, dynamic, flexible, high-risk, scalable business model, innovative
Ripsas et al. (2018)	2018	A Startup Cockpit for the Proof-of-Concept	[...] a young company that is less than 10 years old [...] has an innovative business model and/or deploys innovative technologies. [...] shows significant growth either in the number of employees or in turnover.	young, <10 years old, innovative business model, tech-based, growth-oriented
Kumbhat and Sushil (2018)	2018	Development Stages and Scaling Issues of Startups	[...] living organisms, especially early-stage startups operate under conditions of extreme uncertainty in search of right product–market fit.	faces high uncertainty
Tech (2018)	2018	Financing High-Tech Startups	An organization that aims at scaling revenues and headcount, that is less than 10 years old, and that develops a highly innovative business model or technology.	scale-oriented, <10 years old, innovative business model/ technology
Giardino et al. (2016)	2016	Software Development in Startup Companies: The Greenfield Startup Model	Organizations focused on the creation of high-tech and innovative products, with little or no operating history, aiming to aggressively grow their business in highly scalable markets.	tech-producing, innovative, little/no operating history, growth-/scale-oriented
Thiel and Masters (2014)	2014	Zero to One	New technology tends to come from new ventures – startups [...] a startup is the largest group of people you can convince of a plan to build a different future.	new, innovative, technology developing
Blank and Dorf (2012)	2012	The Startup Owner's Manual	[...] is not a smaller version of a large company. [...] is a temporary organization in search of a scalable, repeatable, profitable business model.	temporary, scalable, seeks innovative business model

(Continued)

Graham (2012)	2012	Startup = Growth	[...] a company designed to grow fast. Being newly founded does not in itself make a company a startup. Nor is it necessary for a startup to work on technology, or take venture funding, or have some sort of "exit." The only essential thing is growth.	growth-oriented, scalable
Ries (2011)	2011	The Lean Startup	[...] a human institution designed to create new products and services under conditions of extreme uncertainty	new, faces high uncertainty
Luger and Koo (2005)	2005	Defining and tracking business start-ups	[...] a business entity: which did not exist before during a given time period (new), which starts hiring at least one paid employee during the given time period (active), and which is neither a subsidiary nor a branch of an existing firm (independent).	new, active, independent

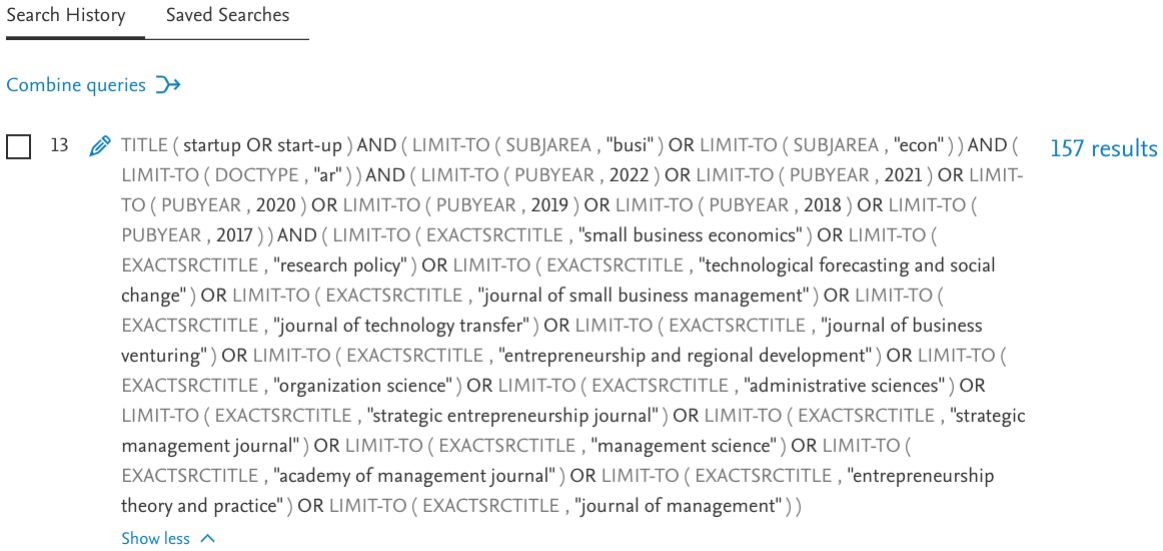


Figure 3: Scopus Query: Research Articles that study and characterize start-ups

VC investment rounds is intended to ensure the suitability of the model for the empirical part of this scientific work.

At the beginning, a start-up is in the early stage, which consists of the pre-seed and seed stage.

In the *pre-seed stage*, there is an initial assumption for a specific problem and the start-up is mainly concerned with finding out whether this problem really exists and if it is meaningful for a large number of prospective customers. The goal of this stage is therefore to identify a problem worth solving and to discover a suitable market for the intended solution (Inc42, 2022; Kumbhat & Sushil, 2018).

This is followed by the *seed stage*, in which the start-up iteratively develops a first deployable solution in collaboration with initial customers and validates whether this solution suf-

Table 5: *Start-up Characteristics in Academic Research*

Characteristic	**Frequency** (#Articles)	**Percentage** (n=18)	**References**
young ($\leq$ 7 years)	7	38.89%	Audretsch et al. (2021), Cacciolatti et al. (2020), Fackler et al. (2022), Roche et al. (2020), Sauermann (2018), Sorenson et al. (2021), and Zahra (2021)
faces high uncertainty	6	33.33%	Audretsch et al. (2021), Cacciolatti et al. (2020), Eveleens et al. (2017), Hasan and Koning (2019), Sorenson et al. (2021), and Zahra (2021)
lack of resources	6	33.33%	Audretsch et al. (2021), Chatterji et al. (2019), S. Lee (2022), Wasserman (2017), Zahra (2021), and Zhang and Gu (2021)
new	6	33.33%	Audretsch et al. (2021), Fackler et al. (2022), Grillitsch and Schubert (2021), S. Lee (2022), Roche et al. (2020), and Sauermann (2018)
small	4	22.22%	Audretsch et al. (2021), Cacciolatti et al. (2020), Fackler et al. (2022), and Sauermann (2018)
develops new products	3	16.67%	Eveleens et al. (2017), Hasan and Koning (2019), and Zhang and Gu (2021)
funded by informal / venture capital	3	16.67%	Cacciolatti et al. (2020), Croce et al. (2018), and Kolokas et al. (2020)
growth-oriented	3	16.67%	Dushnitsky and Yu (2022), Hasan and Koning (2019), and S. Lee (2022)
not directly innovative/ growth-oriented	3	16.67%	Audretsch et al. (2021), Cacciolatti et al. (2020), and S. Lee (2022)
can have any size	2	11,11%	Cacciolatti et al. (2020) and S. Lee (2022)
innovative	2	11,11%	Hasan and Koning (2019) and D. Wang et al. (2022)
vulnerable to economic downturns	2	11,11%	Fackler et al. (2022) and Sorenson et al. (2021)

ficiently satisfies the needs of the target market. This development process is usually associated with the acquisition of initial paying customers and the generation of first revenues. In the next step, the business model is also validated (Blank, 2007; Kumbhat & Sushil, 2018).

If the start-up has the necessary financial resources and the product and business model allow it, it moves into the *growth stage* (Tech, 2018). In this stage, the start-up prepares its processes, product, and business model for rapid growth and subsequently engages in very aggressive customer acquisition to establish a strong market position (Kumbhat & Sushil, 2018; Tech, 2018).

When customer and sales growth slowly level off and sales generated by existing products start to become more and more predictable, the *later stage* is reached. At this stage, VC investors in particular often insist on an exit by means of an IPO or acquisition to realize their profits. To increase sales, a later stage start-up often begins to expand into the international market and/or diversify the product portfolio until at some point the primary focus is on sustaining the company (Inc42, 2022; Tech, 2018).

2.3. Start-up Success in Academic Research

In entrepreneurship and management research, there is no clear consensus on when a start-up can be called successful. Definitions for start-up success are diverse as the meaning of success is always linked to a specific context (Nambisan & Baron, 2013; Santisteban et al., 2021).

2.3.1. Success as the sum of context-specific factors

Maidique and Zirger defined success as “the achievement of something desired, planned, or attempted" (Maidique & Zirger, 1985, p.305). According to this definition, a start-up can be considered successful when it achieves what it was supposed to achieve. The decision as to whether a start-up is successful therefore depends on the expectations placed on it and whether these have been fulfilled (Stuart & Abetti, 1987).

Table 6: *Definitions of a Start-up*

	Early stage **Pre-Seed** (Discovery)	**Seed** (Validation)	**Growth** (Efficiency/Scale)	**Later** (Expansion/Maturity)
Objective	Problem definition and customer discovery	Validation of product and market	Refinement of processes and rapid growth	Expansions on a larger scale and business sustainment
Market / Customers	No real customer, customer discovery through interviews	Market calibration; first paying customers; demand creation	Market penetration; repeatable and aggressive customer acquisition process	Diversification; internationalization; ... market saturation
Product	Value proposition defined, product concept or prototype	First level of deployable solution with core features	Product matured; complementary features added	Mature product offering
Team	Only founders	Founders on full time, few key employees	Rapidly growing	Moderately growing
Revenues	No revenues	Minimal initial revenues	Growing revenues	Stable revenue streams
Funding	Mainly informal	Seed	Series A to B	Series C+, IPO, exit
Investors	Bootstrapping, FFF, BA, crowdfunding	BA, VC, CVC, crowdfunding	VC, CVC, banks	Private / public equity, debt / loans

Note. Adapted from Kumbhat and Sushil (2018) and Tech (2018)

The expectations of a start-up and thus the definition of success depend on a variety of factors.

One of these factors is the perspective from which the start-up is viewed (Nambisan & Baron, 2013). A founder's idea of start-up success, for example, may differ from that of a venture capitalist (Black et al., 2010). For a venture capitalist, a start-up is successful if it generates a high return on investment, i.e., it enables a very profitable exit in the long term – preferably through a successful IPO or acquisition. Success is thus predominantly defined by financial indicators (Davila et al., 2003; Gompers et al., 2008). For a founder, on the other hand, personal and various non-financial criteria are usually relevant in addition to financial aspects when defining success (Fisher et al., 2014).

Another factor that influences the definition of success is a start-up's stage of development. For an early-stage start-up, for instance, success can have a different meaning than for a late-stage startup, as they face different challenges and pursue different milestones (Brush & Vanderwerf, 1992; Chandler & Hanks, 1993; Witt, 2004). While in an early phase the receipt of initial funding can already represent a success, the success in the later stage is rather determined by financial key figures and performance (Witt, 2004; Yua, 2020).

The meaning of success may also depend on the type of start-up (Nambisan & Baron, 2013).

When it comes to sustainable start-ups, success is significantly defined in terms of the contribution to society and the environment, whereas these aspects are usually given less attention in the case of an ordinary venture (Bocken, 2015). The definition of success may also differ depending on the technological complexity of the product offering. While a young software start-up is successful if it achieves a certain growth in sales, for a research-intensive biotechnology start-up, the number of registered patents may be a more appropriate indicator of success (Roche et al., 2020; Vissa & Chacar, 2009).

Lastly, when assessing the success of a start-up, the relevant time horizon is crucial (Nambisan & Baron, 2013). In the short to medium term, rapid growth or the ability to attract rounds of funding may be considered an initial success but is not a reliable indicator of a start-up's long-term success (Stuart & Abetti, 1987). Considering success only in a single time dimension can thus lead to an erroneous assessment (Maltz et al., 2003). Croce et al. (2018) therefore make a distinction between "interim" and "ultimate" success.

2.3.2. Common definitions of success

Success can have different meanings for different individuals, and interpretations can depend on a wide variety of factors (Kiviluoto, 2013; Nambisan & Baron, 2013; Witt, 2004). Nevertheless, there are dimensions in academic research that are more often considered when conceptualizing start-up success than others. Table 7 presents all definitions identified within a sample of 23 scientific articles.

Five definitions of success were particularly prevalent: (1) achieving rapid growth, (2) reaching a high level of financial performance and profitability, (3) attracting various funding rounds, (4) accomplishing a liquidity event such as

Table 7: *Definitions of Start-up Success*

Dimension	Definition	#Articles	References
Growth	Having a highly scalable business model and growing very fast, esp. in terms of sales and employment.	8	Baron and Hannan (2002), Frank et al. (2007), Guo et al. (2021), Hormiga et al. (2011a, 2011b), Peña (2002), Siegel and Wessner (2012), and Stuart and Abetti (1987)
Profitability	High level of financial performance and profitability.	7	Bocken (2015), Hormiga et al. (2011a, 2011b), S. Lee (2022), Roure and Maidique (1986), Stuart and Abetti (1987), and Stucki (2014)
Ability to attract funding	Attracting and closing various funding rounds.	6	Beckman et al. (2007), Gaule (2018), Schlichte et al. (2019), Spiegel et al. (2016), Ter Wal et al. (2016), and Woolley and MacGregor (2021)
Liquidity Event	Accomplishing an IPO or being acquired by another company.	6	Baron and Hannan (2002), Beckman et al. (2007), Croce et al. (2018), Gaule (2018), Hong et al. (2020), and Humphery-Jenner and Suchard (2013)
Survival	Surviving after a defined period of time.	6	Baron and Hannan (2002), Frank et al. (2007), Roure and Maidique (1986), Stuart and Abetti (1987), Stucki (2014), and Woolley and MacGregor (2021)
Goal Attainment	Meeting predefined goals and objectives.	4	Dvir et al. (2010), Maidique and Zirger (1985), L. Z. Song et al. (2010), and Stuart and Abetti (1987)
High Efficiency	Being very efficient in generating output compared to competitors, esp. in the form of revenue or sales.	3	Gloor et al. (2020), L. Z. Song et al. (2010), and Stuart and Abetti (1987)
Positive Founder Perception	Being perceived as successful by the founders across several different aspects.	2	Hormiga et al. (2011a) and Stuart and Abetti (1987)
Contribution	Making a positive contribution to society and the environment.	2	Bocken (2015) and Stuart and Abetti (1987)
Customer / Employee Satisfaction	Satisfying the needs of customers and employees.	1	Stuart and Abetti (1987)

an IPO or acquisition, and (5) surviving after a certain period of time. The diversity of success definitions identified, as well as the fact that 15 of the 23 articles studied (65%) included at least two different dimensions in their definition of startup success, supports the finding of Kiviluoto (2013) and Maltz et al. (2003) that success is a multidimensional construct. This assumption is also the basis of measurement frameworks, such as the "Balanced Scorecard" by Kaplan and Norton (1992) or Dvir et al.'s (1993) "Success Dimensions" (Maltz et al., 2003).

2.3.3. Success vs performance

There is also no uniform definition for the concept of performance in academic literature. Interpretations provided are rarely questioned and an understanding of this construct is usually taken for granted (Richard et al., 2009). Performance was originally defined as the sum of efficiency and effectiveness (Neely et al., 1995). However, performance is increasingly perceived as a multidimensional construct whose concrete meaning also depends on a specific context and thus on a multitude of factors that can change over time (Gerschewski & Xiao, 2015; Kiviluoto, 2013; Laitinen, 2002; Maltz et al., 2003; Richard et al., 2009).

Success and performance are often considered the same in entrepreneurship research because of their seemingly positive relationship to each other (Kiviluoto, 2013; Witt, 2004). This was also evident in the analysis of the studies located in Table 7. In more than half of the 23 reviewed articles, performance was considered an operationalization of success, resulting in the term "performance" often being used as a synonym for "success". Since these concepts can therefore not be considered separately, findings from performance research have been incorporated into this thesis.

2.4. Success Measurement in Academic Research

Success measurement of start-ups is a controversial topic in entrepreneurship and management research (Eveleens et al., 2017; Kiviluoto, 2013; Zahra et al., 2000). This chapter attempts to provide a comprehensive overview of how start-up success has been measured in scientific literature to date and why certain variables, metrics, or measurement methods are utilized. For this purpose, 42 research articles dedicated to the study of start-up success and/or new venture performance were analyzed (see Appendix A3 for a detailed overview of the articles).

2.4.1. Challenges in measuring start-up success

Measuring the success of start-ups is usually associated with special challenges (Eveleens et al., 2017). Obtaining data is mostly difficult, as start-ups do not have to publish financial performance figures and rarely want to share sensitive information with the public (Hope et al., 2013; Schlichte et al., 2019; Woolley & MacGregor, 2021). While it is possible to collect data using self-reporting within surveys, this approach is difficult to implement if very large samples of start-ups are to be studied (Schlichte et al., 2019; Su et al., 2015).

Even if a sufficient amount of data can be collected, there are other aspects to consider when measuring start-up success. Start-ups develop and change very quickly in relatively short periods of time, making the significance of a success measurement highly dependent on the exact point in time at which it was conducted (Eveleens et al., 2017; Garnsey et al., 2006). Moreover, the comparison of measures is often complicated because start-ups can differ significantly depending on their industry and the founders' objectives (Chandler & Hanks, 1993; Eveleens et al., 2017; Witt, 2004). Lastly, studies have shown that performance and success measures of start-ups rarely show significant correlations and can even be negatively correlated (Cooper, 1993; Trailer et al., 1996; Witt, 2004).

2.4.2. Dimensions of success

In analyzing the 42 selected articles, one of the objectives was to identify the specific dimensions that are considered in relevant business research when measuring start-up success.

Frequently observed success dimensions

In examining the selected sample, eight different dimensions of start-up success emerged (see Figure 4).

Most of the measures that were used within the reviewed studies to operationalize success could be assigned to one of these dimensions. The assignment was based on the specification of each study as well as common classifications from entrepreneurship and management literature (Carton & Hofer, 2006; Richard et al., 2009; Trailer et al., 1996). In addition, the determination procedure of a measure or the study-specific data collection method was not considered during assignment. That is, it was not relevant for the classification whether the objective value of a metric or variable was used within a study or whether it was assessed subjectively using a scale. The dimensions most frequently considered in the sample were growth, profitability, and funding. In about 45% of the reviewed studies growth was considered to be a reliable indicator of start-up success. This is in line with the findings of various researchers in the field of entrepreneurship. Growth has been found to be the most used measure of success (Achtenhagen, L. and Naldi, L. and Melin, L., 2010; Davidsson et al., 2009; Kiviluoto, 2013).

Growth as indication of success

The discernible preference for growth as an indication of start-up success may be attributed to the fact that strong business growth is usually associated with above-average results and can have various positive effects on a start-up's development as well as its stakeholders and economy. Growth can not only contribute to the survival of a start-up, but also increase its attractiveness to potential employees and investors (Ben-Hafaïedh & Hamelin, 2022; Coad et al., 2020; Nason & Wiklund, 2018; Phillips & Kirchhoff, 1989). Furthermore, fast-growing companies create a large number of new jobs and are considered a significant driver of a country's economic growth (Coad et al., 2017; Henrekson & Johansson, 2010; Pereira et al., 2020). In addition, business growth is often associated with profitability, as a variety of theories, such as economies of scale, suggest a positive relationship between sales growth and profits (Ben-Hafaïedh & Hamelin, 2022; Nicholls-Nixon, 2005).

However, there is also a growing body of literature that questions this supposedly positive relationship between growth and profitability and warns against viewing growth alone as an indicator of success (Ben-Hafaïedh & Hamelin, 2022; Davidsson et al., 2009; Kiviluoto, 2013). Empirical

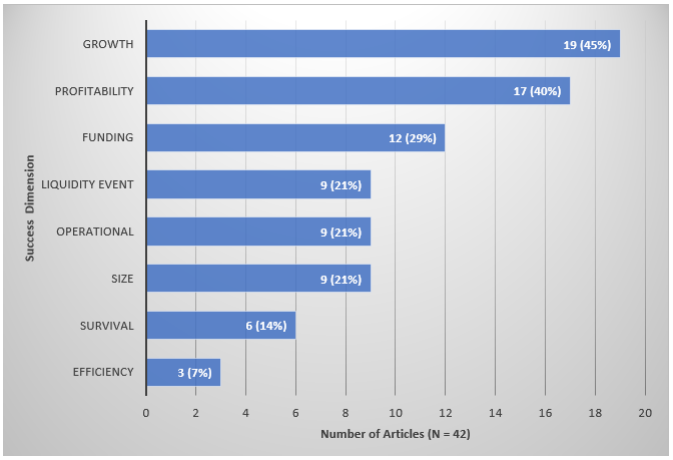


Figure 4: Usage Frequencies of Success Dimensions in Academic Research

Note. Based on Table 8 and Appendix A3, total N = 42 (research articles).

research indicates that growth has at most a weak positive correlation with profitability and may in some cases even be negatively correlated with financial gains (Delmar et al., 2013; Markman & Gartner, 2002; Nason & Wiklund, 2018). Furthermore, studies have shown that at least small and mid-sized companies may increase their likelihood of long-term success by focusing more on profitability than growth in the early years of their operations (Ben-Hafaïedh & Hamelin, 2022; Davidsson et al., 2009). In this regard, Ben-Hafaïedh and Hamelin (2022) state that a strong growth orientation may be appropriate depending on the goals of the entrepreneur and company. They emphasize, however, along with Kiviluoto (2013), that growth alone is rarely a reliable measure of success, and that sustainable growth should be given much greater importance. Growth can thus be of great relevance for the success of start-ups due to their growth-oriented nature, but its causes, reasons as well as its necessity must be questioned (Kiviluoto, 2013).

2.4.3. Types of success measures

In line with the identified dimensions of success, eight distinct types of success measures were detected within the sample studied. These could again be classified into six different categories: growth measures, accounting measures, funding measures, operational measures, liquidity event measures and survival measures. Table 8 provides a complete overview of all observed measures and how they were determined within the 42 empirical studies.

Growth measures

Growth measures comprise all metrics that provide information on the internal and external growth of a start-up (Brinckmann et al., 2011; Carton & Hofer, 2006). Since the long-term economic success of a start-up often depends on how fast it is able to grow, they can be important indicators of success (Grillitsch & Schubert, 2021; Pe'er et al., 2016). Growth measures are considered as a separate category, as they are in some cases clearly distinguished from traditional accounting-based measures in academic literature and growth is usually attributed special importance in the context of measuring the success of start-ups (Grillitsch & Schubert, 2021; Guo et al., 2021). Moreover, in entrepreneurship research, growth is often represented by non-financial indicators (Chatterji et al., 2019; Schlichte et al., 2019).

Among the sampled studies growth was preferably measured by the change in sales or employee numbers. This observation is also consistent with existing literature. According to several researchers, growth in entrepreneurship research is most frequently measured in terms of sales and employment (Achtenhagen, L. and Naldi, L. and Melin, L., 2010; Brinckmann et al., 2011; Grillitsch & Schubert, 2021; Leitch et al., 2010; Witt, 2004). Growth in these two areas is assumed to be a relatively good reflection of the internal and external progress of a start-up (Brinckmann et al., 2011; Delmar & Shane, 2003). Sales growth is a good indicator of how successfully a startup's products or services are accepted and adopted by its target market (Brinckmann et al., 2011; Zahra et al., 2000). Employment growth, on the other hand, is a clear indicator of a startup's overall internal growth and an increase in employee numbers suggests that more capacity is available to realize objectives and generate higher returns for stakeholders (Brinckmann et al., 2011; Gilbert et al., 2006). Despite their ability to capture important developmental aspects of a start-up, both sales and employment growth were rarely used alone in the sample, but rather in combination with other measures of growth to represent this dimension. This reflects the heterogeneity as well as multidimensionality of the growth construct and underlines the difficulty to capture this dimension on the basis of only one best measure (Kiviluoto, 2013; Leitch et al., 2010). Moreover, by using different measures, a more precise understanding of the type of growth achieved can be obtained (Brinckmann et al., 2011).

Table 8: *Measures of Success and Performance in Academic Research*

Dimension	Metrics	Frequency (#Articles)	Type	Determined	References
Growth	Sales growth	12	financial	obj/subj	Brinckmann et al. (2011), Deligianni et al. (2017), Fultz and Hmieleski (2021), Hormiga et al. (2011a, 2011b), C. Lee et al. (2001), Li and Atuahene-Gima (2001), Peña (2002), Ruiz-Jiménez, J. M. and Ruiz-Arroyo, M. and del Mar Fuentes-Fuentes, M. (2021), Stam and Elfring (2008), Vissa and Chacar (2009), and Zahra et al. (2000)
	Employment growth	10	non-financial	obj/subj	Beckman (2006), Brinckmann et al. (2011), Chatterji et al. (2019), Fultz and Hmieleski (2021), Hmieleski et al. (2012), Peña (2002), Ruiz-Jiménez, J. M. and Ruiz-Arroyo, M. and del Mar Fuentes-Fuentes, M. (2021), Schlichte et al. (2019), Siegel and Wessner (2012), and Stam and Elfring (2008)
	Profit growth	4	financial	obj/subj	Deligianni et al. (2017), Hormiga et al. (2011a), Li and Atuahene-Gima (2001), and Peña (2002)
	Market share growth	3	non-financial	obj/subj	Deligianni et al. (2017), Li and Atuahene-Gima (2001), and Zahra et al. (2000)
	Revenue growth	3	financial	objective	Baron and Hannan (2002), Ensley and Hmieleski (2005), and Hmieleski et al. (2012)
	Customer growth	1	non-financial	subjective	Ruiz-Jiménez, J. M. and Ruiz-Arroyo, M. and del Mar Fuentes-Fuentes, M. (2021)
Profitability	Return on investment	5	financial	obj/subj	Hormiga et al. (2011b), Li and Atuahene-Gima (2001), Podoynitsyna et al. (2013), Ruiz-Jiménez, J. M. and Ruiz-Arroyo, M. and del Mar Fuentes-Fuentes, M. (2021), and L. Z. Song et al. (2010)
	Net profit	4	financial	subjective	Chen (2009), Deligianni et al. (2017), Fultz and Hmieleski (2021), and Stam and Elfring (2008)
	Return on assets	4	financial	subjective	Chen (2009), Hormiga et al. (2011a, 2011b), and Li and Atuahene-Gima (2001)
	Return on equity	3	financial	obj/subj	Deligianni et al. (2017), Ruiz-Jiménez, J. M. and Ruiz-Arroyo, M. and del Mar Fuentes-Fuentes, M. (2021), and Zahra et al. (2000)
	Gross profit	2	financial	subjective	Fultz and Hmieleski (2021) and Stam and Elfring (2008)
	EBITDA	1	financial	objective	Cacciolatti et al. (2020)
	Profit break-even (bi)	1	financial	objective	Stucki (2014)
	Return on sales	1	financial	subjective	Li and Atuahene-Gima (2001)

(Continued)

Funding	Total funding amount received	7	financial	objective	Croce et al. (2018), de Mol, E. and Cardon, M. S. and de Jong, B. and Khapova, S. N. and Elfring, T. (2020), Gaule (2018), Gloor et al. (2020), Roche et al. (2020), Schlichte et al. (2019), and Yua (2020)
	Funding received (bi)	5	non-financial	objective	Beckman et al. (2007), Croce et al. (2018), Roche et al. (2020), Santos and Cardon (2019), and Woolley and MacGregor (2021)
	Follow-on funding (bi)	1	non-financial	objective	Croce et al. (2018)
	2nd funding round (bi)	1	non-financial	objective	Ter Wal et al. (2016)
	Time to first VC funding	1	non-financial	objective	Beckman and Burton (2008)
	Number of funding rounds	1	non-financial	objective	Schlichte et al. (2019)
	Liquidity Event Acquisition (bi)	5	non-financial	objective	Gaule (2018), Hong et al. (2020), Humphery-Jenner and Suchard (2013), Roche et al. (2020), and Yua (2020)
	IPO (bi)	5	non-financial	objective	Beckman et al. (2007), Gaule (2018), Hong et al. (2020), Humphery-Jenner and Suchard (2013), and Roche et al. (2020)
	Probability of IPO	2	non-financial	objective	Baron and Hannan (2002) and Croce et al. (2018)
	Time to IPO	2	non-financial	objective	Baron and Hannan (2002) and Beckman and Burton (2008)
Size	Annual sales	11	financial	obj/subj	Chen (2009), Deligianni et al. (2017), Gao et al. (2010), Gloor et al. (2020), Hormiga et al. (2011a, 2011b), S. Lee (2022), Li and Atuahene-Gima (2001), Siegel and Wessner (2012), Söderblom et al. (2015), and M.-C. Wang and Chen (2016)
	Number of employees	1	non-financial	objective	Santos and Cardon (2019)
Survival	Survival (bi)	5	non-financial	objective	Chatterji et al. (2019), Santos and Cardon (2019), Stucki (2014), Woolley and MacGregor (2021), and Yua (2020)
	Probability of survival	1	non-financial	objective	Baron and Hannan (2002)
Efficiency	Asset turnover ratio	1	financial	objective	Gloor et al. (2020)
	Operational efficiency	1	financial	subjective	Li and Atuahene-Gima (2001)
	Sales per employee	1	financial	objective	Hmieleski and Cole (2022)

(Continued)

Operational non-financial & objective	Number of patents	1	non-financial	objective	Roche et al. (2020) and Siegel and Wessner (2012)
	Customer retention rate	1	-	-	Podoynitsyna et al. (2013)
	Number of copyrights	1	-	-	Siegel and Wessner (2012)
	Number of licensing agreements	1	-	-	Siegel and Wessner (2012)
	Number of trademarks	1	-	-	Siegel and Wessner (2012)
	Time to product shipment	1	-	-	Beckman (2006)
	Web traffic	1	-	-	Yua (2020)
non-financial subjective	Market share	3	non-financial	subjective	Chen (2009), Fultz and Hmieleski (2021), and Stam and Elfring (2008)
	Attainment of goals	2	-	-	Hormiga et al. (2011a, 2011b)
	Development speed	2	-	-	Fultz and Hmieleski (2021) and Stam and Elfring (2008)
	Product innovation	2	-	-	Fultz and Hmieleski (2021) and Stam and Elfring (2008)
	Customer satisfaction	1	-	-	Stam and Elfring (2008)
	Product quality	1	-	-	Stam and Elfring (2008)
Other	Business Plan Quality	1	non-financial	subjective	de Mol, E. and Cardon, M. S. and de Jong, B. and Khapova, S. N. and Elfring, T. (2020)
	Cost control	1	financial	subjective	Stam and Elfring (2008)
	Credit rating	1	financial	objective	Cacciolatti et al. (2020)
	Founder satisfaction	1	non-financial	subjective	Hormiga et al. (2011a)
	Meeting economic goals	1	financial	objective	L. Z. Song et al. (2010)
	Net cash flow	1	financial	objective	Ensley and Hmieleski (2005)
	Operating cash flow	1	financial	subjective	Li and Atuahene-Gima (2001)
	Overall reputation	1	non-financial	subjective	Li and Atuahene-Gima (2001)
	Program graduation	1	non-financial	objective	Gao et al. (2010)
	Receipt of government grants (bi)	1	non-financial	objective	Woolley and MacGregor (2021)
	Team performance	1	non-financial	subjective	Santos and Cardon (2019)
	Years of operation	1	non-financial	objective	Santos and Cardon (2019)

Note. References in the rearmost column were underlined if the corresponding metric was subjectively captured in them.

Accounting measures

Accounting measures are one of the most widely used tools for measuring organizational performance and are based primarily on financial information, as found in the financial statements of a company (Carton & Hofer, 2006). Despite their supposedly low informative value for the success of start-ups they were found to be of recognizable relevance within the sample reviewed (Guo et al., 2021). Accounting measures can be further divided into profitability, size as well as efficiency measures (Carton & Hofer, 2006).

Profitability measures include all metrics that provide insights into whether and to what extent a start-up is able to generate financial gains (Carton & Hofer, 2006). Especially in entrepreneurship research, their relevance for measuring the success of a start-up is often debated and questioned, as start-ups are rarely profitable in the first years of operation. Technology-based start-ups in particular usually have to reach a specific size before they can generate their first profits (Guo et al., 2021; C. Lee et al., 2001). Profitability measures that were observed particularly frequently within the reviewed sample were ROI, net profit and ROA. However, these were rarely used as objective values, but were mainly obtained subjectively.

Size measures provide information about the scale of a start-up. Compared to growth measures, these are usually absolute instead of relative or percentage values. Typical size indicators are total sales, the number of employees or the available assets of a start-up.

In empirical research, it is recommended to use size measures mainly as control variables (Carton & Hofer, 2006). However, in the sample studied, total annual sales in particular was comparatively often considered as the dependent variable.

Efficiency measures cover all variables and metrics that demonstrate how well a start-up uses its resources. They often contrast the generated outcome of a start-up such as sales, turnover, or profit with the resources available or deployed. Resources can be both financial and non-financial. They can range from specific assets, such as a start-up's funding or cash to the number of existing employees. Therefore, efficiency measures can also contain non-accounting information that cannot be retrieved from the financial statements of a start-up (Carton & Hofer, 2006; Gloor et al., 2020).

One drawback of efficiency measures is that they are difficult to compare between different start-ups and industries, as efficiency is often defined differently depending on industry and company-specific factors (Carton & Hofer, 2006).

Funding measures

Funding measures include all variables and metrics that provide insight into how capable a start-up is at attracting financial resources from external investors. They mostly refer to funding from venture capital firms as these represent one of the most important outside sources of financing for start-ups (Gloor et al., 2020; Kolokas et al., 2020; Söderblom et al., 2015). Two funding measures that were used extensively across the 42 studies reviewed were total funding amount received and a binary variable indicating whether funding was received at all.

Funding measures are increasingly used to measure start-up success because of the positive implications of external financing for a startup's development and performance. Empirical research shows that start-ups backed by venture capital firms demonstrate significant superiority in areas such as growth, innovation, efficiency, productivity, and product development speed (Chemmanur et al., 2011; Dutta & Folta, 2016; Gloor et al., 2020; Spender et al., 2017). The ability to attract funding from venture capital firms can thus be critical to the survival of a start-up and its long-term success (Alexy et al., 2012; Audretsch et al., 2012; Beckman et al., 2007; Shane & Stuart, 2002). Especially high-tech start-ups usually need multiple rounds of funding within a relatively short period of time for the development as well as commercialization of their products and services (Schlichte et al., 2019; Ter Wal et al., 2016). Funding measures are also commonly applied, as the receipt of funding is considered an important milestone for a start-up, signaling the confidence of external investors and the start-up's progress (Beckman & Burton, 2008; Shane & Stuart, 2002; Woolley & MacGregor, 2021). Moreover, most funding variables can be compared across industries and are accessible online through public websites and company databases such as Crunchbase (de Mol, E. and Cardon, M. S. and de Jong, B. and Khapova, S. N. and Elfring, T., 2020; Ter Wal et al., 2016; Woolley & MacGregor, 2021; Yua, 2020).

Operational measures

Operational measures mainly comprise variables depicting the non-financial outcomes of a start-up and can vary depending on the industry (Carton & Hofer, 2006; Gerschewski & Xiao, 2015). They can range from the satisfaction of specific stakeholders such as customers and employees to concrete market share or productivity measures (Fultz & Hmieleski, 2021; Gerschewski & Xiao, 2015; Hult et al., 2008; Stam & Elfring, 2008). While many of these measures can be objectively captured a large proportion of operational measures are often based on the assessments of selected respondents, such as internal decision-makers of a start-up or certain experts (Carton & Hofer, 2006).

Operational measures are mostly used as a complementation for accounting-based measures because they can capture future opportunities and success potential that has been generated but is not yet evident in financial measures (Gerschewski & Xiao, 2015; Kaplan & Norton, 1992). However, one downside of these measures is that they are rarely comparable across industries and their significance must be viewed critically due to their often difficult-to-quantify nature (Carton & Hofer, 2006).

Liquidity event measures

Liquidity event measures indicate whether or how quickly a start-up was able to liquidate its assets and distribute them to stakeholders as part of a liquidity event such as an IPO or acquisition by another company (Beckman & Burton, 2008; Hong et al., 2020; Roche et al., 2020; Yua, 2020). Alternatively, they can also show how likely a liquidity event is to occur in the future (Croce et al., 2018). Liquidity event measures, which were used particularly often in the sample studied, were two binary variables used to capture whether a start-up experienced an IPO or an acquisition.

Those measures are often chosen as an indicator for final success, as they allow to compare the success of start-ups from different industries and a liquidity event can be an important milestone, especially for start-ups funded with venture capital (Beckman & Burton, 2008; Beckman et al., 2007; Roche et al., 2020).

Survival measures

Survival measures provide information about the survival and survivability of a start-up. Thus, they usually indicate whether a start-up still exists and continues its business activities within a predefined period of time. In this context, the failure of a start-up is usually equated with its closure (Richard et al., 2009; Witt, 2004). Within the sample studied, the most frequent survival measure was a binary variable, denoting whether a start-up survived after a certain time.

Since start-ups usually have to overcome numerous challenges such as liability of newness or smallness, the probability of survival is often very small (Shane, 2009; Soto-Simeone et al., 2020). Technology-based start-ups in particular have comparatively low survival rates, which is why survival can be a meaningful sign that a company is developing positively and can adapt appropriately to the needs of its market (Carton & Hofer, 2006; M. Song et al., 2008). Therefore, survival measures are mostly used as indicators of success, since the survival of a start-up can signal success in other measurement dimensions and, moreover, collecting data for these measures is comparatively easy (Chatterji et al., 2019; Richard et al., 2009; Witt, 2004; Woolley & MacGregor, 2021). These measures, however, are often associated with methodological hurdles in terms of determining an appropriate observation period and distinguishing between different types of company closures (Richard et al., 2009; Witt, 2004; Woolley & MacGregor, 2021). The closure of a start-up does not necessarily mean its failure, as a closure is not always insolvency-related, but can also be the result of a positive event such as a successful acquisition by another company (DeTienne et al., 2015; Fortune & Mitchell, 2012; Richard et al., 2009; Wennberg et al., 2010). Therefore, a distinction must be made between positive and negative closure (Woolley & MacGregor, 2021). Complementing this, Trailer et al. (1996) note that survival does not always mean success, as a venture can continue to exist despite poor performance.

2.4.4. Measurement preferences

To obtain a more holistic overview of success measurement in the scientific literature, the exact measurement behavior across the selected 42 articles was investigated. The corresponding results are presented in this chapter.

Variety of measures and dimensions of success

According to Gerschewski and Xiao (2015), a combination of different types of metrics from different measurement dimensions should be used to measure performance. This multi-dimensional measurement approach has also been recommended by a variety of other researchers in the field of performance and success measurement (Hult et al., 2008; Richard et al., 2009; Stam & Elfring, 2008; Wiklund & Shepherd, 2005). Within the sample studied, an average of three measures were used to evaluate success (see Figure 5). About 60% of the articles used three or fewer of the measures listed in Table 8.

Figure 6 presents how often a certain number of dimensions was examined in the sample.

About 69% of the sample considered a maximum of two different dimensions when measuring success. One third (33%) focused on only a single dimension. The highest number of measures and dimensions were observed in studies that used subjective measures of success (Fultz & Hmieleski, 2021; Hormiga et al., 2011a).

Subjective and objective measures

In performance research, a distinction is made between objective and subjective measures (Richard et al., 2009; Ruiz-Jiménez, J. M. and Ruiz-Arroyo, M. and del Mar Fuentes-Fuentes, M., 2021; Stam & Elfring, 2008). Subjective measures are determined with the help of subjective assessments by selected respondents. Objective measures, in contrast, originate from objective sources, such as, for example, accounting systems, financial statements or alternatively self-reports within surveys (Richard et al., 2009). Subjective measures are particularly useful for examining non-financial dimensions of start-up success and offer more flexibility than objective measures (Richard et al., 2009; Ruiz-Jiménez, J. M. and Ruiz-Arroyo, M. and del Mar Fuentes-Fuentes, M., 2021; Stam & Elfring, 2008). Empirical studies have shown that subjective measures correlate significantly with objective ones and thus can be considered sound measures when used appropriately (Fultz & Hmieleski, 2021; Wall et al., 2004). However, they can be more susceptible to bias and error and therefore subjective measures are often supplemented by objective measures (Deligianni et al., 2017; Li & Atuahene-Gima, 2001; Richard et al., 2009). Within the sample studied, a preference for objective measures was evident (see Figure 7).

About 74% of the 42 reviewed articles exclusively deployed objective measures for measuring success. Only eight

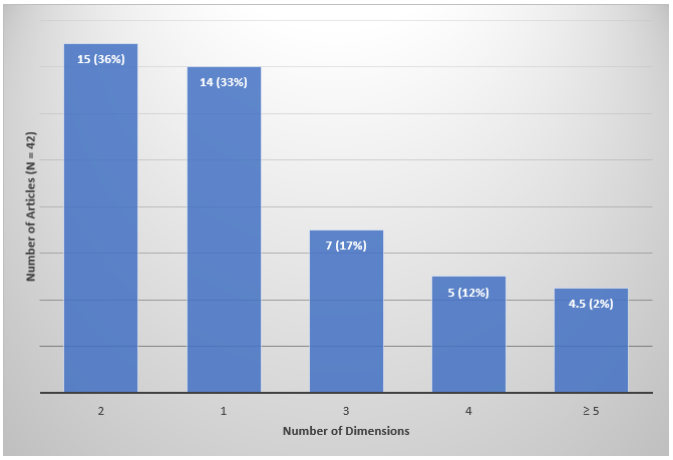


Figure 5: Number of success dimensions considered in empirical Studies

Note. Based on Table 8 and Appendix A3, total N = 42 (research articles).

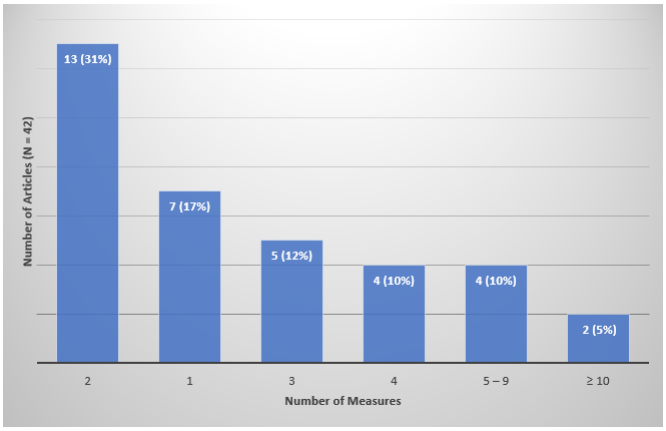


Figure 6: Number of used Success Metrics in empirical Studies

Note. Based on Table 8 and Appendix A3, total N = 42 (research articles).

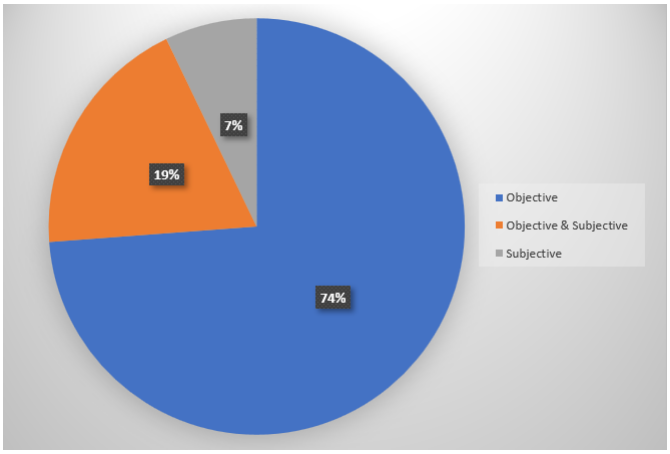


Figure 7: Usage of Subjective and Objective Measures in empirical Studies

Note. Based on Table 8 and Appendix A3, total N = 42 (research articles).

studies (19%) used a combination of objective and subjective measures.

Financial and non-financial measures

Financial measures such as accounting metrics are often used as a measurement tool to assess whether a company has been able to achieve its economic goals (Gerschewski & Xiao, 2015). However, it is increasingly recommended that these are complemented by non-financial measures such as operational metrics, since financial measures alone often fail to capture important components of success (Gerschewski & Xiao, 2015; Kaplan & Norton, 1992; Laitinen, 2002). Figure 8 illustrates the frequency of use of financial and non-financial measures in the sample studied.

The majority of the sample (62%) used a combination of financial and non-financial measures to depict success. However, still more than one fifth of the sample (21%) examined success only at the financial level.

2.4.5. Success measurement across different start-up stages

The definition of success may differ depending on the stage of a start-up's development. That is there can also be differences in how success is measured across different development stages (Brush & Vanderwerf, 1992; Chandler & Hanks, 1993; Witt, 2004). Start-ups in a very early stage, for example, rarely have a product that can be commercialized and therefore usually do not generate any revenue or sales. Financial metrics are therefore not suitable for measuring the success of such ventures. In general, it is usually hard to capture the success of a very young early-stage start-up using quantitative metrics (de Mol, E. and Cardon, M. S. and de Jong, B. and Khapova, S. N. and Elfring, T., 2020; Witt, 2004). According to the explanations of Witt (2004) and the measurement decisions observed in the 42 selected articles, all identified measures as well as measure types were assigned to the four development phases of a start-up (see Figure 9).

3. Empirical Research – Practitioners' Perspective

This study examines how start-up success is defined and measured among venture capitalists. By doing so, it aims to capture the practitioners' perspective on the central research question to complement the academics' view and gain a more comprehensive understanding of how start-up success is measured. The primary focus of the study is specifically on very young early-stage start-ups as these were only marginally represented in the literature review. However, success measurement in this phase represents an important object of investigation for this thesis since success in the early days of a start-up is comparatively difficult to measure (de Mol, E. and Cardon, M. S. and de Jong, B. and Khapova, S. N. and Elfring, T., 2020; Witt, 2004).

3.1. Research Methodology

A qualitative research design was used to gain an in-depth understanding of the views and opinions of the selected practitioners (Patton, 2002). Specifically, semi-structured expert interviews were employed to obtain information-rich insights, and to complement the literature review previously conducted (Flick, 2009).

3.1.1. Participant Sampling

Within a start-up ecosystem, there is a wide range of different participants in diverse roles, who moreover tend to have varying areas of responsibility and objectives (Cukier & Kon, 2018; Kiviluoto, 2013; Tripathi et al., 2019). It is hardly feasible to sufficiently capture the views of all relevant groups of practitioners in such an ecosystem. For this reason, this study focuses exclusively on one group of practitioners that was deemed to be particularly important within a start-up ecosystem: venture capitalists. This group of practitioners was chosen because the venture capital field can have a significant impact on the success of start-ups and venture capitalists may have valuable expert knowledge about the topic of success measurement gained from working with diverse types of start-ups (Alexy et al., 2012; Beckman et al., 2007; Gloor et al., 2020; Hall & Hofer, 1993; Maula et al., 2005). It has therefore been assumed that venture capitalists, because of their extensive experience with and knowledge of start-ups, can be considered experts whose knowledge, despite their mostly financial intentions, can potentially compensate for the lack of other practitioners' perspectives (Davila et al., 2003).

To determine an appropriate sample of respondents with a high degree of knowledge about the topic of interest, a multi-stage sampling process was conducted. First, based on the explanations of Patton (2002), purposeful sampling, specifically criterion sampling was used to identify a sample of 22 early-stage venture capital funds. For this, the LinkedIn database was used. The selection of VC companies was based on three different criteria. To be included in the sample, a VC firm had to (1) be headquartered in Europe or the UK, (2) have a special focus on for-profit early-stage technology start-ups respectively pre-seed and seed investments and (3) have a minimum number of 15 start-ups in its portfolio. Subsequently, based on three criteria, more than 57 experts were identified within the selected VC companies and contacted via LinkedIn. According to (Singh, 2015), an expert should have proven domain knowledge and experience in the relevant research domain. To be considered an expert, a potential interview participant therefore had to (1) work at a VC fund in an investment-driven position, (2) work primarily with early-stage start-ups, and (3) be active in venture capital for at least 1.5 years and/or demonstrate comparable experience with start-ups. The sampling and contacting process resulted in a final sample of seven experts from seven different early-stage VC funds with an average VC experience of more than three years.

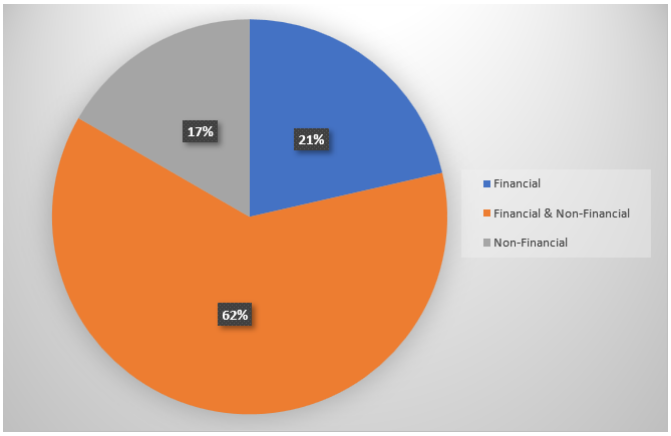


Figure 8: Usage of Financial and Non-financial Measures in empirical Studies

Note. Based on Table 8 and Appendix A3, total N = 42 (research articles).

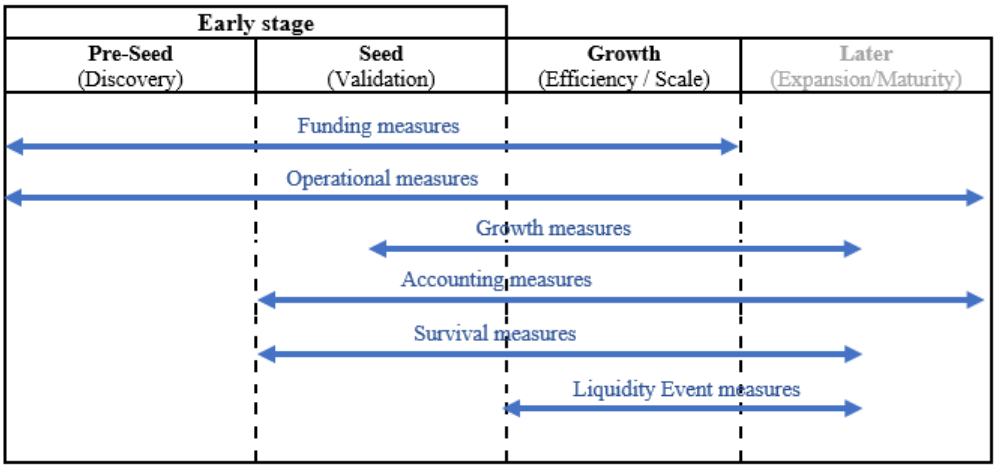


Figure 9: Success Measures across Development Stages

Note. Based on Table 8 and Appendix A3, later stage start-ups were hardly represented within the reviewed studies.

3.1.2. Data Collection

To gain a better understanding of the phenomenon of start-up success and how it is measured, seven semi-structured expert interviews were conducted with seven selected early-stage venture capitalists. An overview of these is provided in Table 9. The interviews were carried during the period of June 21-29, 2022, using the video-chatting service Google Meet, and lasted between 19 and 28 minutes, with an average duration of approximately 24 minutes. Additionally, a two- to three-minute introduction to the research was given prior to each interview.

It was decided to use semi-structured interviews to be able to elaborate on the respondents' statements and to give them the opportunity to shape the conversation themselves. This was intended to contribute positively to the substantive scope of the interviews (Singh, 2015). To allow for more consistency between interviews and completeness of content, an interview guide with key questions was created (see Appendix A4). This guide was based on the academic literature previously reviewed and was validated beforehand through a test interview with a venture capitalist from the author's network. Furthermore, the interview guide was divided into three consecutive phases, whereby in each phase it was ensured that open-ended questions were used to minimize biases (Singh, 2015). In the first phase, an attempt was made to obtain an overview of the respondent and its areas of responsibility. In the second phase, the expert was asked to conceptualize the concepts of "start-up" and "start-up success". The definitions of these key constructs were then used as a transition for the third phase, which specifically focused on success measurement. If requested, participants were provided with the prepared interview guide prior to the interview. The interviews conducted were audio recorded using software integrated into the author's laptop. Each respondent approved this recording with a prior signature of a con-

Table 9: *Overview of Interviewed Experts*

Expert	Expert's role	Investment focus	Interview duration	Interview date
E1	Investment Analyst	pre-seed	23:52	21.06.2022
E2	Investment Associate	pre-seed	24:57	23.06.2022
E3	Managing Director	early-stage	19:03	23.06.2022
E4	Investment Manager	pre-seed to seed	24:04	24.06.2022
E5	Senior Associate	early-stage	28:35	24.06.2022
E6	Investment Analyst	early-stage	23:05	24.06.2022
E7	Partner	early to growth stage	24:33	29.06.2022

firmation of consent as well as a second verbal confirmation before the start of the interview.

3.1.3. Data Analysis

After the data collection was completed, a thematic analysis was conducted following the phases and specifications of Braun and Clarke (2006). Table 10 (Braun & Clarke, 2006, p.87) illustrates the specific phases that were followed.

First, the audio recordings of each interview were reviewed and transcribed verbatim (see Appendix B for entire interview transcripts). Specifically, an intelligent verbatim transcription was performed to make the spoken text more readable. Thereby, all types of identifiers such as names, specific job titles, and the like were omitted to ensure the anonymity of interview participants. Finally, all interviews conducted in German were translated into English by using the neural machine translation software DeepL. To distinguish the individual interviews from each other in the further process, each expert was assigned a unique identifier (E1-E7). After the transcription, the interview data were coded using the data analysis software ATLAS.ti. Codes were defined inductively and then assigned to predefined as well as inductively determined themes. The final coding scheme is presented in Table 11.

3.2. Venture Capitalists' View of Success

3.2.1. Definitions of start-up success

At the beginning of the interview, the study participants were asked what they considered to be a start-up and when they considered a start-up to be successful. Start-ups were primarily described as young, innovative growth companies that often have a specific technology focus and are usually VC funded (see Appendix A5). To define a successful start-up, some of the respondents incorporated different perspectives and emphasized the different meanings of success to different viewers. Thus, three definitions of success emerged as seen in Table 12.

Contributing to the fund's performance: Four venture capitalists defined success in terms of their fund's financial goals. They described a successful start-up as one that contributes significantly to the fund's return. For example, one respondent stated: "*With the VC fund, it is relatively clear to achieve high valuations quickly.* (E3)". This statement was confirmed by another interviewee, again highlighting the importance of high valuations for the fund's viability and continued existence:

> *Speaking as a VC, the start-up has to be a unicorn. I mean we have 99 percent write-offs until we pay back our fund. [...] If we get our money back a hundred or two hundred times with a company, then that's enough to pay back the entire fund.* (E2)

Creating added value and generating profits: By taking a less investment-oriented perspective, some of the respondents came to the conclusion that a start-up is successful if it creates real value for customers while achieving a good financial performance and eventually profitability. One respondent pointed out that this should be the ultimate goal of a start-up:

> *[...] that the start-up manages to translate this growth potential into a sustainably functioning business model and bring innovative solutions to the market that customers can then benefit from. That would be the goal for me. That's how I would define success*. (E1)

Proving the viability of the company: Two venture capitalists further considered it a success if a start-up can overcome certain stage dependent hurdles and reach its milestones, proving that it is viable:

> *The way I think about things is maybe: "What is meaningful signal for a company to prove risks of that company that it's viable?" [...] But I think it comes back to just really thinking about what are the big milestones to prove as a company and where they are that in that journey [...].* (E5)

Table 10: *Overview of Interviewed Experts*

Phase	Description of the process
1. Familiarizing yourself with your data	Transcribing data (if necessary), reading and re-reading the data, noting down initial ideas.
2. Generating initial codes	Coding interesting features of the data in a systematic fashion across the entire data set, collating data relevant to each code.
3. Searching for themes	Coding interesting features of the data in a systematic fashion across the entire data set, collating data relevant to each code.
4. Reviewing themes	Checking if the themes work in relation to the coded extracts (Level 1) and the entire data set (Level 2), generating a thematic 'map' of the analysis.
5. Defining and naming themes	Checking if the themes work in relation to the coded extracts (Level 1) and the entire data set (Level 2), generating a thematic 'map' of the analysis.
6. Producing the report	The final opportunity for analysis. Selection of vivid, compelling extract examples, final analysis of selected extracts, relating back of the analysis to the research question and literature, producing a scholarly report of the analysis.

Note. From Braun and Clarke (2006)

Table 11: *Interview Coding Scheme*

First-order codes	Second-order themes	Aggregated Thematic Domains
• Contributing to the fund's performance • Creating added value and generating profits • Proving the viability of the company	Definition of start-up success	VCs' view of start-up success
• Being able to attract good investors • Getting closer to product-market fit • Making rapid progress in the right direction	Definition of start-up success in the early stage	
• Success as multi-dimensional construct • Success as context-specific construct	Success as a complex construct	
• Development stage • Business model • Product • Business type	Measurement influencing factors	Measuring start-up success in early stages
• Direct statements • Indirect statements	Lack of data	
• Traction • Product & engagement • Team & organization • External Validation • Other measures	Preference for operational metrics	
• Traction metrics	Traction metrics	
• Product metrics • Engagement metrics	Product & engagement metrics	
• Arguments in favor of growth measures • Arguments against growth measures	Rapid growth	Assessing the significance of growth & funding
• Arguments in favor of funding measures • Arguments against funding measures	Receipt of equity funding	

Table 12: *Respondents' Definitions of Start-up Success*

Success is identified as	**Respondents' Statements**
Contributing to the fund's performance	• *For us, from a pre-seed fund standpoint, I would say success is classified as, "You manage to raise a strong next round." [...] Then later on, again, the question will come whether you go for an IPO or whether you try to sell that or sell individual shares.* (E1) • *Speaking as a VC, the start-up has to be a unicorn. I mean we have 99 percent write-offs until we pay back our fund. You can do that if a start-up really increases its value at least a hundredfold from our point of view or increases our value in the company a hundredfold. If we get our money back a hundred or two hundred times with a company, then that's enough to pay back the entire fund.* (E2) • *It is different from the accelerator to the VC fund. With the VC fund, it is relatively clear to achieve high valuations quickly.* (E3) • *From an investor's perspective, it is often said that a start-up is successful as soon as it raises a lot of venture capital.* (E6)
Creating added value and generating profits	• *[...] that the start-up manages to translate this growth potential into a sustainably functioning business model and bring innovative solutions to the market that customers can then benefit from. That would be the goal for me. That's how I would define success.* (E1) • *We are a financial investor, which means we naturally look at the figures and then you can define success on the basis of traction, sales, revenue, growth. [...]Only if the product is good and successful, creates added value and has a market demand, then that naturally results in positive financial KPIs, which are then again defined with success. [. . .] or a financial investor, success is really defined by revenue and growth [. . .].* (E4) • *[. . .] but for me to really call a company successful, it either has to have a very clear path to profitability that I find credible or actually have already achieved profitability.* (E6)
Proving the viability of the business	• *The way I think about things is maybe: "What is meaningful signal for a company to prove risks of that company that it's viable?" [...] But I think it comes back to just really thinking about what are the big milestones to prove as a company and where they are that in that journey [...]* (E5) • *I think a successful start-up is a company that is young, as I said, and relatively new to the market with a new business model, with a new product, with a new service, whatever, and has proven that there is a product market fit. And, that the company is growing.* (E7)

3.2.2. Success in the early stage

After explaining what they understood by a successful start-up, the interview participants were asked for their definition of success in the case of an early-stage start-up. Although respondents had a relatively similar understanding of what an early-stage start-up is (see Appendix A6), a definition was previously provided to ensure a consistent understanding of this term. Again, three major definitions of success became apparent as seen in Table 13.

Being able to attract good investors: One theme that came up frequently in the context of early-stage success was raising additional funding. Three of the respondents considered it a clear success if a start-up was able to raise further rounds of financing from attractive investors following their initial investment. Respondent E1 made this particularly clear:

> *For us, from a pre-seed fund standpoint, I would say success is classified as, "You manage to raise a strong next round." [...] for us, really, the most important thing is that they manage to raise a strong follow-on round. [...] For really early stage, pre-seed start-ups, I would say very clearly: a strong follow-on round with prominent investors.* (E1)

Getting closer to product-market fit: Some respondents considered it a particular early-stage success when a start-up shows early signs of a so-called product-market fit. This was seen as a clear indication that the company is developing positively:

> *In principle, whenever you have at least five customers who are really convinced of your solution. [. . .] then you can perhaps also, although you still have few customers and are still at an early stage,*

Table 13: *Respondents' Definitions of early-stage Start-up Success*

Success is defined as	**Respondents' Statements**
Being able to attract good investors	• *For us, from a pre-seed fund standpoint, I would say success is classified as, "You manage to raise a strong next round." [...] for us, really, the most important thing is that they manage to raise a strong follow-on round. [...] For really early stage, pre-seed start-ups, I would say very clearly: a strong follow-on round with prominent investors.* (E1) • *It's all very individual. What's also a success criterion is, are you getting relevant investors? Do you have the ability to sell that well? Do people believe in you? That's also super important to measure success. For our start-ups, the external validation. Does anyone else believe in that? If so, how good are they, how much value do they bring?* (E2) • *It's great if they manage to achieve a high valuation or a higher valuation, and if they still manage to attract attractive investors.* (E3)
Getting closer to product-market fit	• *I think in the early stage, it's successful if you can see that they're moving in the right direction. [...] Are you getting closer to this product market fit? That is interconnected with the metrics that you map at the end. Do they grow and do you get a benchmark somewhere where you can say, "this is getting closer and closer to a product market fit".* (E2) • *[...] it gets back to that question of like, what's the signal of the company? What are the actual key risks? And so, you know, as I said, if you look at a company, and you might look at a company. The risk here is that even if you build it, it won't be valuable. And so, if they're in that part of the journey, where they've got to a stage where they have built it, and they've got it out in the market, and people are using it, and you know that it's not exciting people and retention numbers aren't going well, then they probably haven't proven the market risk attribute level. Like we're not even getting to business risks, yet. They haven't been able to prove the market risk. Maybe they've proven the product risk, they could build product. But the market risk is now kind of gone.* (E5) • *In principle, whenever you have at least five customers who are really convinced of your solution. There is the classic product-driven question "How bad would you feel or how unsatisfied would you be if our product were no longer available to you overnight?" and if a certain percentage then really says "on a scale of 1-10, it's an 8" then you can perhaps also, although you still have few customers and are still at an early stage, very strongly assume that it is a very good product and they must be close to product market fit. And accordingly, that would be one of the main success metrics.* (E6)
Making rapid progress in the right direction	• *I think in the early stage, it's successful if you can see that they're moving in the right direction. Now that's a super big question because it's a case-by-case decision. I think if you see the team manages to develop the product very quickly in one direction and add more and more features and be more and more responsive to customer needs, I think that's relatively important on the product side. If they can iterate quickly, execute quickly. Then, on the traction side, we want to see that it's going in the right direction, that they're gaining more and more customers, that they're staying longer, that they're slowly paying for it.* (E2) • *[...] they are successful...if they are naturally in our portfolio, we assume for the time being that they will continue to grow and can act positively in the market environment. [...] Success is when the team is well established, when a good organization is built up after our investment, [...] a team that is resistant to such downs when things don't go so well. They have a strong team, they can successfully deal with those problems, find solutions and move on. [...] and if they have a sales structure, that they have a strong funnel, that they can successfully convert customers from first approach into paying customers. [...] The product has to be successful. That has to offer the features and the services and solutions that are promised, that we've invested in. [...] Revenue growth over month above 10 percent is always successful for me. On the other hand, it can also be successful if it responds to market events and has to pivoted if necessary, product has to be adapted. That's what I meant at the beginning with "resistant to adversities that occur" [...].* (E4)

(Continued)

	• *I think you look very much in phases [...] in which you set milestones [...] Where you look at okay, where are we today? What is our plan? Where do we want to go? And then you look, are we achieving these goals, yes or no? [...] And I think that's when you realize you're on a good path. And honestly, in an early phase, goals often change, because you go to the market with hypotheses. It's a very iterative process. Sometimes you realize okay, the hypothesis that we have set up is maybe not quite right, maybe we need to tweak the product a little bit, adjust it so that we get in the direction of product market fit. [...] That is, you may no longer be running exactly in the direction that you set as your goal six months ago, but you are still on the right path [...] that's how I would define success. [...] Are you getting closer to your goals in the planned horizon [...]* (E7)

> *very strongly assume that it is a very good product, and they must be close to product market fit.* (E6)

Consistent with E6's statement, E7 considered an increasing use of the product as a meaningful sign of product-market fit. Furthermore, E7 defined product-market-fit as a state in which there is a pull effect from the market, i.e., customers actively demand a startup's products without the company having to draw their attention to them.

Making rapid progress in the right direction: Furthermore, a very young start-up was considered successful if it makes rapid progress and develops positively in various business dimensions. Particular relevance was attributed to the product and customer side and the team:

> *I think if you see the team manages to develop the product very quickly in one direction [...] I think that's relatively important on the product side. [...] we want to see that it's going in the right direction, that they're gaining more and more customers, that they're staying longer, that they're slowly paying for it.* (E2)

In this regard, E7 stated that the progress of a start-up can be primarily assessed by whether the planned milestones have been reached within a specified time horizon. However, E7 also emphasized that these milestones can change very quickly in the early stage.

3.2.3. Success as a complex construct

As respondents engaged in conceptualizing success, it became apparent that they perceived success as a complex, multiform construct that must be considered on multiple dimensions and whose meaning depends on contextual factors (see Appendix A7). For example, one venture capitalist stated: *"Anyone that tries to encapsulate success in one metric is wrong [...] if you're taking one in isolation from these attributes, I think you, you can't get a full picture. (E5)"*.

3.3. Measuring Start-up Success in early stages

3.3.1. Measurement influencing factors

In answering questions about success measurement, four of the interview participants pointed out that the specific metrics choice can vary depending on certain factors. The primary factors mentioned were business type, business model, development stage and the product of the start-up (see Appendix A8). This was again summarized by respondent E7:

> *So, to identify the golden KPI or North Star metric for an early-stage start-up is, I think, difficult, because it depends very much on the business model, is it a B2B or B2C company, the product and the phase.* (E7)

3.3.2. Preference for operational metrics

The respondents' statements highlighted that in the early stages of a start-up's development, there is little to no historical or financial data available to measure success (see Appendix A9). The majority of the venture capitalists therefore emphasized the need for operational metrics:

> *Basically, I think we can probably say that the goals by which one would ultimately define success in the early phase are not so much aimed at financial KPIs, but rather at operational KPIs or other milestones. The older the company, the greater the importance of the financial KPIs.* (E7)

E7 also highlighted that product milestones are particularly important for young start-ups, as the pre-seed and seed phases are primarily about product development and product improvement: *"I'm thinking here of pre-seed, seed companies. It's very much about developing a product, finding product market fit, which then at some point also means having the first customers who use the product. (E7)"*. However, some respondents noted that the early phases are very fluid (see Appendix A10) and there can be a rather quick transition from these operational metrics and product milestones to common metrics such as revenue growth:

> *I think these are the first points that you have to validate so that you are somehow successful and as soon as you have found something there, it's about these revenue metrics. [...] But as soon as [...] that slowly moves in a direction of a growing company, the metrics MRR and revenues are super important. That's a fluid stage [...].* (E2)

Table 14 shows all metrics mentioned in the interviews.

Table 14: *Success Metrics in early Stages*

Metrics	**Respondents' Statements**
Traction	
Number of customers	• *So how much interest is there on the user side? How many customers do they have in the pipeline?* (E1) • *How is your pipeline evolving? [. . .] Then, on the traction side, we want to see that it's going in the right direction, that they're gaining more and more customers [. . .]* (E2) • *[...] That's where I think it's more relevant to say, "have you been able to find paying customers? Yes or no?", [. . .] And then maybe the number and how well you can convert them.* (E6)
Paying customers (bi)	• *Then, on the traction side, we want to see [. . .] that they're slowly paying for it. [. . .]and that you test the willingness to pay. That they are willing to pay a certain price for it.* (E2) • *[...] That's where I think it's more relevant to say, "have you been able to find paying customers? Yes or no?" [. . .]* (E6) • *Usage, so do I have customers who use the product? And ideally, they are also willing to put money on the table.* (E7)
Number of pilot / PoC / development customers	• *I think it's important that you have one, two, three customers that you develop your product with. So the number of POCs is definitely important.* (E2) • *Pilot projects are very helpful. That is, of course, a very good indicator. [...] if they win customers, and paid pilot projects, I always think that's very important. That simply says a lot..* (E3) • *I think of very, very early-stage [...] it is simply more important to find a development customer who wants to use the product and is willing to pay something. [. . .] the first question is "do you find development customers and are they willing to pay for it?"* (E6)
WTP	• *Then, on the traction side, we want to see [. . .] that they're slowly paying for it. [. . .]and that you test the willingness to pay. That they are willing to pay a certain price for it.* (E2) • *[. . .] We also look at willingness to pay, which is also very important, but the first question is "do you find development customers and are they willing to pay for it?"* (E6)
PoC size	• *So the number of POCs is definitely important. How big are the POCs? How big are the departments that are testing the product?* (E2)
Traction vs age	• *There's a dimension that we call "traction versus age." That's always an absolute gut decision, but we try to score how much traction they have now compared to other companies that we see in the portfolio compared to how long they've been in the game.* (E2)
Number of people interested in product	• *Non-financial measures, for example, if you do a closed beta, how many people want to get on your platform? How many people want to use that?* (E1)
Length of sales cycle	• *I think sales cycle is also ultra important in a B2B SaaS. If they get shorter that's always good as well.* (E2)
First CAC	• *Maybe it's not sales and revenue, but it's [. . .] how the first customer acquisition costs are going.* (E4)
Product & Engagement	
Customer Churn	• *And then already the validation. The customers don't churn right away, I think that's important.[. . .] and look at "When are these customers churning? Why are they churning?".* (E2) • *[. . .] churn among customers is of course incredibly important.* (E3) • *And if there is not this churn, it means of course first of all that the product can establish itself or is used.* (E4)

(Continued)

Product development speed	• *And how can the team iterate with the feedback? How quickly can they develop the product? I think that's important. [. . .] even if you can't quantify it, how fast paced are they? [. . .] all that matters to us is growth, speed, and in the beginning you don't have metrics that grow, but can you manage to test your product quickly, iterate. Are you fast there?* (E2) • *product growth. The speed at which iteration takes place, so to speak.* (E6) • *How many sprints can we manage to achieve the goal?* (E7)
Retention rate	• *[. . .] but more non-financial metrics that show engagement. Also, retention rate and things like that. [. . .] How often do they come back?* (E1) • *Then, on the traction side, we want to see that it's going in the right direction, that they're gaining more and more customers, that they're staying longer, that they're slowly paying for it. [. . .] Also super important.[. . .] What is your seven day retention? What is your 30 day retention?* (E2)
Stickiness	• *Otherwise, it's interesting to measure stickiness from a product perspective, for example, using the retention rate.* (E1)
Session length	• *When people are on the platform, how long do they stay? How much time do they spend there?* (E1)
Active users	• *Also super important. How many one active users? How many daily active users? How many weekly active users?* (E2)
Net promoter Score	• *[. . .] and if a certain percentage then really says "on a scale of 1-10, it's an 8" then you can perhaps [. . .] very strongly assume that it is a very good product and they must be close to product market fit. And accordingly, that would be one of the main success metrics.* (E6)
Achievement of product milestones	• *We have defined milestones at product level where we want to get to.* (E7)
Product quality	• *And how well do we deliver on the tech side, on the specifications from the product?* (E7)
Team & Organization	
Team churn	• *It's also important for the team. If you notice that people are always leaving, that's also a huge issue.* (E3) • *What is their employee turnover? How many people leave the company?* (E6)
Achievement of team building or organizational goals	• *[. . .] there are also goals that we usually have that are very strong in terms of team building and organizational development.* (E7)
Team quality	• *Maybe it's not sales and revenue, but it's [. . .] how you build the team and how the first customer acquisition costs are going.* (E4) • *And to qualitatively still people and hiring. Who do they hire? How good are the people they herd, but also the people they have?* (E6)
External Validation	
Ability to attract good follow-on investors (bi)	• *For us, from a pre-seed fund standpoint, I would say success is classified as, "You manage to raise a strong next round." [...] And I would say that very clearly, for pre-seed early stage, a strong follow-on round is actually the most relevant performance indicator.* (E1) • *What's also a success criterion is, are you getting relevant investors?[. . .]* (E2)
Attractiveness for network	• *Then, of course, also attractiveness for the network.* (E3)
Others	
Business plan quality	• *Of course, that's always less numbers driven with pre-seed. [. . .] where we just go really deep is if they can just get a good business plan.* (E3)
Burn rate / runway	• *I think a good indicator of success is also how long or how well they can plan ahead. Start-ups that have to raise three rounds because they didn't manage to pay attention to their runway in the first round is always difficult.* (E3)

3.3.3. Traction metrics

In most of the interviews, one indicator of success was mentioned particularly often: Traction.

Consistent with the statements of the other respondents, E1 defined traction as evidence that users are interested in a start-up's product:

> *In the early stage, it's also always very important for us to see, [...] but simply traction. So how much interest is there on the user side? How many customers do they have in the pipeline? Things like that. traction.* (E1)

Moreover, E2 attributed great relevance to the first signs of traction for early success, as they provide insight into how close a start-up is to potential product-market fit:

> *I think to make this a bit more tangible, the combination of traction compared to age is relatively important, i.e. how quickly do you manage to build up traction, even if these are only the early signs of traction. [...] Are you getting closer to this product market fit?* (E2)

A variety of different metrics were used to measure traction (see Table 14). The most frequently stated indicator was the number of (paying) customers a start-up is able to attract to co-develop the product or conduct a pilot project or proof of concept.

3.3.4. Product and engagement metrics

Besides traction, respondents placed particular emphasis on metrics that provide insight into product progress and user engagement with the product (see Table 14). One product-related measure respondents perceived as very critical was the speed at which a start-up develops its solution. In this regard, E2 argued that even in the early stages, measuring the speed and growth of a startup is essential, but common metrics are rarely available:

> *How quickly can they develop the product? I think that's important. [...] even if you can't quantify it, how fast paced are they? [...] all that matters to us is growth, speed, and in the beginning you don't have metrics that grow, but can you manage to test your product quickly, iterate. Are you fast there?* (E2)

Furthermore, two engagement metrics that came up most often during the interviews were customer churn and user retention. E4 considered the churn rate to be a good indicator that a product is being used sufficiently and is able to succeed in its market.

3.4. Assessing the Significance of Growth & Funding

In academic literature, success has often been measured using growth or funding measures (see Chapter 2.4.2 and 2.4.3). Interview participants were therefore asked for their opinions on the use of these metrics as indicators of success, especially in the early stages of a start-up.

3.4.1. Rapid growth

Although the majority of respondents attributed great importance to growth, only the minority perceived rapid growth as a significant indicator of success as seen in Table 15.

Especially in the early stages, typical growth measures were considered a less meaningful indicator because most very young start-ups are not ready to scale. Furthermore, E1 and E3 attributed significantly more importance to sustainable growth and pointed out that, regardless of the phase, growth alone is not a sufficient measure of success:

> *Just bottomless growth without looking at the costs on the same side is, of course, dangerous. A classic is: The revenue shoots up, but in the same proportion, the costs shoot up, and so you can't build a sustainable, successful business model.* (E1)

3.4.2. Receipt of equity funding

While two respondents considered the receipt of a strong funding round as a clear success, this

potential success indicator was also viewed rather critically as seen in Table 16.

On the one hand, some significance was attached to funding measures, as external financing can be necessary to achieve fast growth , may be an indication of good start-up performance and has the potential to attract further attractive investors. On the other hand, four respondents highlighted the fact that funding is only a means to an end, and it is better to be able to grow quickly even with less equity funding:

> *But I don't think the funding itself is a success at this point. It's actually much better if you can grow quickly with less money. [...] That brings less dilution, more efficient growth, faster profitability, and that's also something that is generally very highly valued by the market.* (E7)

4. Discussion and Implications

Both in research and in practice, special attention is always paid to one particular type of start-ups: successful ones. To date, however, little research has been done on how to measure a start-up's success (Eveleens et al., 2017; Kiviluoto, 2013). This thesis therefore examined what academics as well as practitioners understand by a successful start-up and what they consider to be reliable measures of success. In addition, it outlined the rationale for certain measurement decisions.

The literature analysis shows that researchers from the entrepreneurship and management fields measure the success of a start-up primarily by its growth, profitability, and/or ability to raise rounds of financing. Especially growth and funding measures are considered significant indicators of success, as a large number of empirical studies confirm their relevance for the performance of a start-up (Gloor et al., 2020). However, a handful of authors question the use

Table 15: *Opinions on Growth as an Indication of Success*

Rapid growth ...	**Respondents' Statements**
is success indicator / success	• *[...] even if you can't quantify it, how fast paced are they? At the end of the day, all that matters to us is growth, speed, and in the beginning you don't have metrics that grow, but can you manage to test your product quickly, iterate. Are you fast there? Do you have the right nose? [...] I think that is the main indicator. You just have to look in which dimension growth. Growth in the team. Growth in founder development. Growth in product maturity. Growth in terms of the traction KPIs, but at the end of the day it's all about growth in some dimension.* (E2) • *I think for a start-up and for a financial investor, success is really defined by revenue and growth, especially if the two are related. [. . .] Sales growth is crucial for me. Sales or revenue. That is of course what is important. In the beginning, it's more sales and traction to customers.* (E4) • *Yes, absolutely. In the end, it's always about growth. [. . .] So, financing as fuel, fuel as we say, to accelerate growth, to drive it forward. [...] I would say, it's mre of a means to an end to achieve your actual goals, your actual success* (E7)
to be questioned	• *Of course, it always depends a bit on the business model. What are they trying to sell? Just bottomless growth without looking at the costs on the same side is, of course, dangerous. A classic is: The revenue shoots up, but in the same proportion, the costs shoot up, and so you can't build a sustainable, successful business model. In general, in addition to these growth metrics, I always find it interesting to see whether a product market fit has been achieved in a certain way because you can read off quite well from that or that signals relatively strongly whether sustainable, profitable growth is possible or whether you are simply forced to pump more and more money into marketing in order to keep sales high.* (E1) • *I would rather say sustainable growth. Growth alone is out of the question. Because we have seen that often enough, that companies are taken over very quickly and that then only the figures are polished for the investors. That's why early stage is so difficult, because sustainable growth is usually not yet in there. [...] I don't necessarily believe in rapid growth, but rather in sustainable growth.* (E3) • *There are start-ups where it can be a good indicator. [...]But it's not necessarily like that. When we invest in pre-seed, it's often the case that there are a lot of pilot customers, but that's not converting yet, so they don't have the big sales yet. [...] Then there are other KPIs that are paid attention to. Maybe it's not sales and revenue, but it's the customers you talk to, how you build the team and how the first customer acquisition costs are going. And that's important again on the other side. Sales is an important indicator, but it can also take a back seat in the very early stages.* (E4) • *And you could say it's successful, depending on maybe looking at growth metrics, and a bunch of things. But then you look at a space company, or a biology company or a semiconductor company. The signal for those companies might take a lot more money and a lot more time.* (E5) • *Generally speaking, no. It depends very much on whether it's a business where you know that a lot of human capital will be needed, because it's still a very operations-heavy business. There growth is extremely important [...] That would be growth of people level. If it's a pure B2B SaaS business, it's quite important how many customers can you onboard and sell, so to speak, and then also revenue growth, of course. [...] It is important, but I can't tell you, if it's just about growth in general, which level I would always look at first.* (E6) • *But even in the early stage, there are milestones that make up success before you just get into the growth scaling phase. Probably more product level milestones, organizational milestones, so team building, recruiting relevant positions, roles, product milestones.* (E7)

of growth as a measure of success and state that sustainable growth is a much better indicator of long-term success (Ben-Hafaïedh & Hamelin, 2022). In line with Eveleens et al.'s (2017) suggestion, the reviewed literature also indi-

Table 16: *Arguments for and against Funding as an Indicator of Success*

	Respondents' Statements
Pro	
Enables growth	• *I think that's one of the most valuable indicators.[...] I think funding is definitely a success indicator. If you don't have money, you can't grow.* (E2) • *There are different business models, quick commerce and everything similar just burns an insane amount of money. Of course, you need a different funding level.* (E3) • *[...] Growth stories always mean burning a lot of money.* (E4) • *When it comes to profitability, raising funding allows you to grow much faster [...]* (E6) • *[...] at the end of the day, growth usually costs money somewhere, of course, and especially in the venture area, where it's about rapid growth, about scaling, you need cash [...] So, financing as fuel, fuel as we say, to accelerate growth, to drive it forward. [...] I would say, it's more of a means to an end to achieve your actual goals, your actual success. In that sense, it's also a building block somewhere, a piece of the puzzle of the overall success.* (E7)
Suggests very good performance	• *I would say that it is particularly important for early stage start-ups. Especially because many other KPIs that you would otherwise use are simply not available. [...] And then for us, the first validation that we see is, okay, they manage to raise a strong next round, then that's definitely the most important success characteristic in the stage.* (E1) • *That's a very valuable indicator because you can assess so quickly, what are cases that hit anything. Nerve of the time, good timing, a good market, good product, good team. If they get good investors and raise a stable round, so to speak, then that's the indicator "this can be something".* (E2) • *The investment rounds are a clear indicator of where the company stands. Seed investments usually go up to a maximum of 3 million. [...] From this you can deduce how roughly the company is valued [...] you're in a rage mostly so up to 10 million pre month. That's early and you can actually deduce from that where such a company stands. These are sales of up to one million recurring per year, because in the software sector you can be significantly higher. And if this much revenue is generated, you can deduce approximately, not exactly of course, how many customers there are, how far along the product is and whether the product still needs to be developed further. That's why round size is the decisive factor for me.* (E4) • *So used to series A or Series B very, very hard to raise, and there wasn't that many people in that market, you know, there was 10 firms on Sand Hill Road in SF, and you know, if you got through, that was a real test, and you actually must have proven something.* (E5)
Potentially attracts good follow-on investors (positive signaling)	• *So this signalling is a very important thing. Which investor makes the deal? There's this notion that this is a magnetic industry. Good people track good people. Good investors, good founders, good team, etc. If you have good investors, then again they track the best follow-on investors.* (E2) • *[...] it is a reality that funding can help you a lot in terms of publicity, and then a self-reinforcing circle can form. You have been raised by good investors, you get more publicity. You will also probably find better follow-up investors with a higher likelihood, because it is simply a fact that when investors invest, it is always checked very carefully who invested before What portfolio companies do they have in that area?". Accordingly, yes, relevant.* (E6)
Brings in valuable advisors & partners	• *[...] I think in the early stage it's more important where the money comes from, how much support you get there [...]* (E2) • *But the ideal way is that every 12-24 months a financing round comes along that fits into the equity story, that is significantly better, that more great investors and partners come in that bring added value, in example internationalization or other topics that help [...]* (E4)

(Continued)

Contra	
Lower funding needs may also be due to higher efficiency or more thoughtful funding decisions	• *[...] the important thing is whether you have the right funding for the business model. So rather that the funding matched with what you need. I don't think the amount of funding alone is relevant. [...] But that has to fit your business model, that you don't go back to fundraising every six months, but that there is also a bit of a runway.* (E3) • *Of course, there might be times when you need to raise money really quickly and then you need to do the next round, but in the early stage, funding rounds are a pain for the founding team. [...] and when we have invested, we first want to see that they work with the money and then ideally only raise the next round in 12-15 months. After all, they should be able to handle the money 18-24 months.* (E4)
Being profitable and growing fast with less funding is better	• *When it comes to profitability, raising funding allows you to grow much faster, but I have a thousand times more respect for companies that manage to grow and give away very little equity.* (E6) • *But I don't think the funding itself is a success at this point. It's actually much better if you can grow quickly with less money. [...] That brings less dilution, more efficient growth, faster profitability, and that's also something that is generally very highly valued by the market. You can see that right now in the last six months where the markets have cooled down a bit. Profitability is very much appreciated by investors in particular on markets. [...] if I'm profitable [...] or the more profitable I am, the less external financing needs I usually have.* (E7)
Cannot be considered in isolation	• *You can't look at it in isolation. Either you're a killer fundraiser to get good follow-on funding because you have a super network and you're a good salesperson. Or you have a good growth in I don't know what dimensions so that somebody new goes in there.* (E2) • *I would say, it's more of a means to an end to achieve your actual goals, your actual success. In that sense, it's also a building block somewhere, a piece of the puzzle of the overall success. [...] But I wouldn't consider a financing round in itself, taken in isolation, to be a success. [...]* (E7)
Informative value of financing decisions must be questioned	• *But it's not now I mean [...] in Europe, we went from five, six years ago, we had about 600 fundraising rounds in a year, and now we're up to about 4000. So, a lot more, a lot more investors a lot more rounds. So, I think there's less signal in the fundraising to the success of the company. I think sometimes actually it can be the wrong thing. People get funded too early because it's a hype play. And often they then don't focus on the fundamentals of the company. [...] So, if you are getting rewarded for not having great metrics, so in terms of this revenue multiple in the growth model you don't need to focus on people giving you money.* (E5)

cates that a combination of different types of measures is preferable for measuring the success of start-ups due to the multidimensional nature of the success phenomenon (Gerschewski & Xiao, 2015; Richard et al., 2009). It also became apparent, however, that academics face specific challenges in measuring start-up success that may limit their choice of success measures (Eveleens et al., 2017).

The qualitative study reveals that venture capitalists, despite their financial intentions, perceive start-up success as a multidimensional construct whose concrete definition depends on a wide variety of factors. How they specifically measure success therefore depends for example on the business model, development phase, or product of a start-up. According to the venture capitalists interviewed, reliable financial metrics are usually not available in the very early stages of a start-up. Operational metrics and milestones at the product or organizational level were therefore considered the preferred options for measuring initial success. Along with this, the interviews once again confirmed that growth and the ability to raise rounds of equity financing can be important indicators of start-up success, but that their necessity and cause should always be questioned.

4.1. Theoretical Implications

This scientific work contributes to a more holistic and better understanding of the success phenomenon in start-up research. The results show substantial overlaps with the findings of Kiviluoto (2013): Start-up success has multiple dimensions that need to be considered and measuring it using only one metric or dimension is rarely recommended. According to the experts interviewed, even growth measures should be viewed critically (Ben-Hafaïedh & Hamelin, 2022; Davidsson et al., 2009). This questions the validity of studies that have used only one indicator to measure success (S.

Lee, 2022; Söderblom et al., 2015). The limited informative value for start-up success attributed to funding measures also contradicts the measurement approach of studies that have exclusively used measures of this type (Ter Wal et al., 2016). Furthermore, the findings of the empirical study support Nambisan and Baron's (2013) as well as Witt's (2004) statement that success and its measurement depend on various factors, such as perspective or development stage.

4.2. Practical Implications

At a practical level, the results of the study are valuable for both researchers and practitioners as they present how the success of start-ups can be measured and what needs to be considered. The findings from the literature review as well as the empirical research suggest that start-up success is a multi-dimensional phenomenon that can hardly be measured by only one metric or captured within one single dimension. Scholars and practitioners could therefore benefit from a more differentiated view and measurement of start-up success that specifically takes into account its multi-dimensionality. In particular, the literature reviewed indicates that financial, non-financial, subjective as well as objective measures of different success dimensions should ideally be combined in future research to provide a more meaningful measurement of start-up success. Furthermore, venture capitalists are provided with opinions and experiences from peers in their field, allowing them to gain deeper insights into what to look for when measuring start-up success, especially in the early stages of development. Other relevant practitioners can especially use the results of the empirical research to better understand how venture capitalists think about and measure start-up success. Founders, for example, can thus better understand the expectations of their investors both before and after raising VC funding.

4.3. Limitations and Future Research

Although the goals of this academic work could be achieved, this was still only possible to a limited extent. Start-ups and their success have been the focus of academic research for some time, but very limited research has been conducted on the topic of success measurement. Moreover, in entrepreneurship and management research, an explicit distinction between success and performance is rarely made. Therefore, in the course of the literature review, the available success literature had to be supplemented by findings from performance research. Within the empirical research, only the perspective of venture capitalists was examined, leaving other relevant practitioners of a start-up ecosystem, such as founders, unrepresented in the results. In addition, despite the strong commitment to scientific rigor, the results of the empirical study are hardly generalizable due to the qualitative approach and the relatively small sample size. To address these limitations, future research could examine the findings of the empirical study using a large-scale quantitative approach or replicate the study including significantly more types of relevant practitioners. Future research could also attempt to develop a standardized success measurement framework for technology start-ups taking into account, for example, the measurement challenges of researchers

5. Conclusion

This thesis aimed to advance the knowledge about success measurement in start-up research by providing a comprehensive overview of what academics as well as practitioners define as a successful start-up and what they consider to be reliable measures of success. To achieve this goal, several scientific studies dedicated to the investigation of start-up success were analyzed and semi-structured expert interviews with venture capitalists from the early-stage investment sector were conducted.

It was found that start-up success in the academic world is particularly measured by the growth of a start-up, its profitability and/or ability to raise external capital. The empirical research revealed that venture capitalists mostly prefer common financial metrics for measuring start-up success. However, for very early-stage start-ups they rely mainly on operational metrics as well as product and organizational milestones due to the lack of financial data. Both the academic and practical perspectives suggest that start-up success is a complex, context-specific phenomenon whose measurement depends on a variety of different factors that may change over time. A meaningful measurement of start-up success therefore requires the use of a combination of different types of metrics to address this complex nature of success. To determine a start-up's success, for example, by considering only one metric or dimension is a dangerous approach that does not account for the true character of the success construct.

References

Achtenhagen, L. and Naldi, L. and Melin, L. (2010). "Business growth" - Do practitioners and scholars really talk about the same thing? *Entrepreneurship: Theory and Practice*, *34*(2). https://doi.org/10.1111/j.1540-6520.2010.00376.x

Alexy, O. T., Block, J. H., Sandner, P., & ter Wal, A. L. J. (2012). Social capital of venture capitalists and start-up funding. *Small Business Economics*, *39*(4), 835–851. https://doi.org/10.1007/s11187-011-9337-4

Audretsch, D. B., Belitski, M., & Caiazza, R. (2021). Start-ups, Innovation and Knowledge Spillovers. *Journal of Technology Transfer*, *46*(6), 1995–2016. https://doi.org/10.1007/s10961-021-09846-5

Audretsch, D. B., Bönte, W., & Mahagaonkar, P. (2012). Financial signaling by innovative nascent ventures: The relevance of patents and prototypes. *Research Policy*, *41*(8), 1407–1421. https://doi.org/10.1016/j.respol.2012.02.003

Baron, J. N., & Hannan, M. T. (2002). Organizational blueprints for success in high-tech start-ups: Lessons from the Stanford Project on Emerging Companies. *California Management Review*, *44*(3), 8–36. https://doi.org/10.2307/41166130

Beckman, C. M. (2006). The influence of founding team company affiliations on firm behavior. *Academy of Management Journal*, *49*(4), 741–758. https://doi.org/10.5465/AMJ.2006.22083030

Beckman, C. M., & Burton, M. D. (2008). Founding the future: Path dependence in the evolution of top management teams from Founding to IPO. *Organization Science*, *19*(1), 3–24. https://doi.org/10.1287/orsc.1070.0311

Beckman, C. M., Burton, M. D., & O'Reilly, C. (2007). Early teams: The impact of team demography on VC financing and going public. *Journal of Business Venturing*, *22*(2), 147–173. https://doi.org/10.1016/j.jbusvent.2006.02.001

Ben-Hafaïedh, C., & Hamelin, A. (2022). Questioning the Growth Dogma: A Replication Study. *Entrepreneurship: Theory and Practice*. https://doi.org/10.1177/10422587211059991

Black, E. L., Burton, F. G., Wood, D. A., & Zimbelman, A. F. (2010). Entrepreneurial Success: Differing Perceptions of Entrepreneurs and Venture Capitalists. *The International Journal of Entrepreneurship and Innovation*, *11*(3), 189–198. https://doi.org/10.5367/000000010792217272

Blank, S. G. (2007). *The Four Steps to the Epiphany: Successful Strategies for Products that Win* (3rd). Quad/Graphics.

Blank, S. G., & Dorf, B. (2012). *The Startup Owner's Manual* (1st). K&S Ranch.

Bocken, N. M. P. (2015). Sustainable venture capital - Catalyst for sustainable start-up success? *Journal of Cleaner Production*, *108*, 647–658. https://doi.org/10.1016/j.jclepro.2015.05.079

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brinckmann, J., Salomo, S., & Gemuenden, H. G. (2011). Financial Management Competence of Founding Teams and Growth of New Technology-Based Firms. *Entrepreneurship: Theory and Practice*, *35*(2), 217–243. https://doi.org/10.1111/j.1540-6520.2009.00362.x

Brush, C. G., & Vanderwerf, P. A. (1992). A comparison of methods and sources for obtaining estimates of new venture performance. *Journal of Business Venturing*, *7*(2), 157–170. https://doi.org/10.1016/0883-9026(92)90010-O

Cacciolatti, L., Rosli, A., Ruiz-Alba, J. L., & Chang, J. (2020). Strategic alliances and firm performance in startups with a social mission. *Journal of Business Research*, *106*, 106–117. https://doi.org/10.1016/j.jbusres.2019.08.047

Carton, R. B., & Hofer, C. W. (2006). *Measuring organizational performance: Metrics for entrepreneurship and strategic management research*. Edward Elgar.

Chandler, G. N., & Hanks, S. H. (1993). Measuring the performance of emerging businesses: A validation study. *Journal of Business Venturing*, *8*(5), 391–408. https://doi.org/10.1016/0883-9026(93)90021-V

Chatterji, A., Delecourt, S., Hasan, S., & Koning, R. (2019). When does advice impact startup performance? *Strategic Management Journal*, *40*(3), 331–356. https://doi.org/10.1002/smj.2987

Chemmanur, T. J., Krishnan, K., & Nandy, D. K. (2011). How does venture capital financing improve efficiency in private firms? A look beneath the surface. *Review of Financial Studies*, *24*(12), 4037–4090. https://doi.org/10.1093/rfs/hhr096

Chen, C.-J. (2009). Technology commercialization, incubator and venture capital, and new venture performance. *Journal of Business Research*, *62*(1), 93–103. https://doi.org/10.1016/j.jbusres.2008.01.003

Christensen, C. M., & Bower, J. L. (1996). Customer power, strategic investment, and the failure of leading firms. *Strategic Management Journal*, *17*(3), 197–218. https://doi.org/10.1002/(sici)1097-0266(199603)17:3<197::aid-smj804>3.0.co;2-u

Coad, A., Cowling, M., & Siepel, J. (2017). Growth processes of high-growth firms as a four-dimensional chicken and egg. *Industrial and Corporate Change*, *26*(4), 537–554. https://doi.org/10.1093/icc/dtw040

Coad, A., Frankish, J. S., & Storey, D. J. (2020). Too fast to live? Effects of growth on survival across the growth distribution. *Journal of Small Business Management*, *58*(3), 544–571. https://doi.org/10.1080/00472778.2019.1662265

Cooper, A. C. (1993). Challenges in predicting new firm performance. *Journal of Business Venturing*, *8*(3), 241–253. https://doi.org/10.1016/0883-9026(93)90030-9

Croce, A., Guerini, M., & Ughetto, E. (2018). Angel Financing and the Performance of High-Tech Start-Ups. *Journal of Small Business Management*, *56*(2), 208–228. https://doi.org/10.1111/jsbm.12250

Crossan, M. M., & Apaydin, M. (2010). A multi-dimensional framework of organizational innovation: A systematic review of the literature. *Journal of Management Studies*, *47*(6), 1154–1191. https://doi.org/10.1111/j.1467-6486.2009.00880.x

Cukier, D., & Kon, F. (2018). A maturity model for software startup ecosystems. *Journal of Innovation and Entrepreneurship*, *7*(1), Article 14. https://doi.org/10.1186/s13731-018-0091-6

Davidsson, P., Steffens, P., & Fitzsimmons, J. (2009). Growing profitable or growing from profits: Putting the horse in front of the cart? *Journal of Business Venturing*, *24*(4), 388–406. https://doi.org/10.1016/j.jbusvent.2008.04.003

Davila, A., Foster, G., & Gupta, M. (2003). Venture capital financing and the growth of startup firms. *Journal of Business Venturing*, *18*(6), 689–708. https://doi.org/10.1016/S0883-9026(02)00127-1

de Mol, E. and Cardon, M. S. and de Jong, B. and Khapova, S. N. and Elfring, T. (2020). Entrepreneurial passion diversity in new venture teams: An empirical examination of short- and long-term performance implications. *Journal of Business Venturing*, *35*(4), 105965. https://doi.org/10.1016/j.jbusvent.2019.105965

Deligianni, I., Voudouris, I., & Lioukas, S. (2017). Do Effectuation Processes Shape the Relationship Between Product Diversification and Performance in New Ventures? *Entrepreneurship: Theory and Practice*, *41*(3), 349–377. https://doi.org/10.1111/etap.12210

Delmar, F., McKelvie, A., & Wennberg, K. (2013). Untangling the relationships among growth, profitability and survival in new firms. *Technovation*, *33*(8–9), 276–291. https://doi.org/10.1016/j.technovation.2013.02.003

Delmar, F., & Shane, S. (2003). Does business planning facilitate the development of new ventures? *Strategic Management Journal*, *24*(12), 1165–1185. https://doi.org/10.1002/smj.349

DeTienne, D. R., McKelvie, A., & Chandler, G. N. (2015). Making sense of entrepreneurial exit strategies: A typology and test. *Journal of Business Venturing*, *30*(2), 255–272. https://doi.org/10.1016/j.jbusvent.2014.07.007

Dushnitsky, G., & Yu, L. (2022). Why do incumbents fund startups? A study of the antecedents of corporate venture capital in China. *Research Policy*, *51*(3), 104463. https://doi.org/10.1016/j.respol.2021.104463

Dutta, S., & Folta, T. B. (2016). A comparison of the effect of angels and venture capitalists on innovation and value creation. *Journal of Business Venturing*, *31*(1), 39–54. https://doi.org/10.1016/j.jbusvent.2015.08.003

Dvir, D., Sadeh, A., & Malach-Pines, A. (2010). The fit between entrepreneurs' personalities and the profile of the ventures they manage and business success: An exploratory study. *Journal of High Technology Management Research*, *21*(1), 43–51. https://doi.org/10.1016/j.hitech.2010.02.006

Dvir, D., Segev, E., & Shenhar, A. (1993). Technology's varying impact on the success of strategic business units within the miles and snow typology. *Strategic Management Journal*, *14*(2), 155–161. https://doi.org/10.1002/smj.4250140206

Ensley, M. D., & Hmieleski, K. M. (2005). A comparative study of new venture top management team composition, dynamics and performance between university-based and independent start-ups. *Research Policy*, *34*(7), 1091–1105. https://doi.org/10.1016/j.respol.2005.05.008

Eveleens, C. P., van Rijnsoever, F. J., & Niesten, E. M. I. (2017). How network-based incubation helps start-up performance: a systematic review against the background of management theories. *Journal of Technology Transfer*, *42*(3), 676–713. https://doi.org/10.1007/s10961-016-9510-7

Fackler, D., Hölscher, L., Schnabel, C., & Weyh, A. (2022). Does working at a start-up pay off? *Small Business Economics*, *58*(4), 2211–2233. https://doi.org/10.1007/s11187-021-00508-2

Fisher, R., Maritz, A., & Lobo, A. (2014). Evaluating entrepreneurs' perception of success: Development of a measurement scale. *International Journal of Entrepreneurial Behaviour and Research*, *20*(5), 478–492. https://doi.org/10.1108/IJEBR-10-2013-0157

Flick, U. (2009). *An introduction to qualitative research: Theory method and applications* (4th). SAGE Publications.

Fortune, A., & Mitchell, W. (2012). Unpacking firm exit at the firm and industry levels: The adaptation and selection of firm capabilities. *Strategic Management Journal*, *33*(7), 794–819. https://doi.org/10.1002/smj.972

Frank, H., Lueger, M., & Korunka, C. (2007). The significance of personality in business start-up intentions, start-up rialization and business success. *Entrepreneurship and Regional Development*, *19*(3), 227–251. https://doi.org/10.1080/08985620701218387

Fultz, A. E. F., & Hmieleski, K. M. (2021). The art of discovering and exploiting unexpected opportunities: The roles of organizational improvisation and serendipity in new venture performance. *Journal of Business Venturing*, *36*(4), 106121. https://doi.org/10.1016/j.jbusvent.2021.106121

Gao, J., Li, J., Cheng, Y., & Shi, S. (2010). Impact of initial conditions on new venture success: A longitudinal study of new technology-based firms. *International Journal of Innovation Management*, *14*(1), 41–56. https://doi.org/10.1142/S1363919610002544

Garnsey, E., Stam, E., & Heffernan, P. (2006). New firm growth: Exploring processes and paths. *Industry and Innovation*, *13*(1), 1–20. https://doi.org/10.1080/13662710500513367

Gaule, P. (2018). Patents and the Success of Venture-Capital Backed Startups: Using Examiner Assignment to Estimate Causal Effects. *Journal of Industrial Economics*, *66*(2), 350–376. https://doi.org/10.1111/joie.12168

Gerschewski, S., & Xiao, S. S. (2015). Beyond financial indicators: an assessment of the measurement of performance for international new ventures. *International Business Review*, *24*(4), 615–629. https://doi.org/10.1016/j.ibusrev.2014.11.003

Giardino, C., Paternoster, N., Unterkalmsteiner, M., Gorschek, T., & Abrahamsson, P. (2016). Software Development in Startup Companies: The Greenfield Startup Model. *IEEE Transactions on Software Engineering*, *42*(6), 585–604. https://doi.org/10.1109/TSE.2015.2509970

Gilbert, B. A., McDougall, P. P., & Audretsch, D. B. (2006). New venture growth: A review and extension. *Journal of Management*, *32*(6), 926–950. https://doi.org/10.1177/0149206306293860

Gloor, P. A., Fronzetti Colladon, A., Grippa, F., Hadley, B. M., & Woerner, S. (2020). The impact of social media presence and board member composition on new venture success: Evidences from VC-backed U.S. startups. *Technological Forecasting and Social Change*, *157*, 120098. https://doi.org/10.1016/j.techfore.2020.120098

Gompers, P., Kovner, A., Lerner, J., & Scharfstein, D. (2008). Venture capital investment cycles: The impact of public markets. *Journal of Financial Economics*, *87*(1), 1–23. https://doi.org/10.1016/j.jfineco.2006.12.002

Graham, P. (2012). *Startup = Growth*. http://www.paulgraham.com/growth.html

Grillitsch, M., & Schubert, T. (2021). Does the timing of integrating new skills affect start-up growth? *Strategic Entrepreneurship Journal*, *15*(4), 647–684. https://doi.org/10.1002/sej.1375

Guo, H., Yang, J., & Han, J. (2021). The Fit between Value Proposition Innovation and Technological Innovation in the Digital Environment: Implications for the Performance of Startups. *IEEE Transactions on Engineering Management*, *68*(3), 797–809. https://doi.org/10.1109/TEM.2019.2918931

Hall, J., & Hofer, C. W. (1993). Venture capitalists' decision criteria in new venture evaluation. *Journal of Business Venturing*, *8*(1), 49–58. https://doi.org/10.1016/0883-9026(93)90009-T

Hasan, S., & Koning, R. (2019). Prior ties and the limits of peer effects on startup team performance. *Strategic Management Journal*, *40*(9), 1394–1416. https://doi.org/10.1002/smj.3032

Henrekson, M., & Johansson, D. (2010). Gazelles as job creators: a survey and interpretation of the evidence. *Small Business Economics*, *35*, 227–244.

Hmieleski, K. M., & Cole, M. S. (2022). Laughing All the Way to the Bank: The Joint Roles of Shared Coping Humor and Entrepreneurial Team-Efficacy in New Venture Performance [in press]. *Entrepreneurship: Theory and Practice*. https://doi.org/10.1177/10422587211046543

Hmieleski, K. M., Cole, M. S., & Baron, R. A. (2012). Shared Authentic Leadership and New Venture Performance. *Journal of Management*, *38*(5), 1476–1499. https://doi.org/10.1177/0149206311415419

Hong, S., Serfes, K., & Thiele, V. (2020). Competition in the venture capital market and the success of startup companies: Theory and evidence. *Journal of Economics & Management Strategy*, *29*(4), 741–791. https://doi.org/10.1111/jems.12394

Hope, O.-K., Thomas, W. B., & Vyas, D. (2013). Financial reporting quality of U.S. private and public firms. *Accounting Review*, *88*(5), 1715–1742. https://doi.org/10.2308/accr-50494

Hormiga, E., Batista-Canino, R. M., & Sánchez-Medina, A. (2011a). The Impact of Relational Capital on the Success of New Business Start-Ups. *Journal of Small Business Management*, *49*(4), 617–638. https://doi.org/10.1111/j.1540-627X.2011.00339.x

Hormiga, E., Batista-Canino, R. M., & Sánchez-Medina, A. (2011b). The role of intellectual capital in the success of new ventures. *International Entrepreneurship and Management Journal*, *7*(1), 71–92. https://doi.org/10.1007/s11365-010-0139-y

Hult, G. T. M., Ketchen, D. J., Griffith, D. A., Chabowski, B. R., Hamman, M. K., Dykes, B. J., Pollitte, W. A., & Cavusgil, S. T. (2008). An assessment of the measurement of performance in international business research. *Journal of International Business Studies*, *39*(6), 1064–1080. https://doi.org/10.1057/palgrave.jibs.8400398

Humphery-Jenner, M., & Suchard, J.-A. (2013). Foreign VCs and venture success: Evidence from China. *Journal of Corporate Finance*, *21*(1), 16–35. https://doi.org/10.1016/j.jcorpfin.2013.01.003

Inc42. (2022). *What Are the Different Stages Of A Startup?* https://inc42.com/resources/what-are-the-different-stages-of-a-startup/

Jin, L., Madison, K., Kraiczy, N. D., Kellermanns, F. W., Crook, T. R., & Xi, J. (2017). Entrepreneurial Team Composition Characteristics and New Venture Performance: A Meta-Analysis. *Entrepreneurship: Theory and Practice*, *41*(5), 743–771. https://doi.org/10.1111/etap.12232

Kaplan, R. S., & Norton, D. P. (1992). The balanced scorecard: Measures That drive performance. *Harvard Business Review*, *83*(7–8), 71–79.

Kazanjian, R. K. (1988). Relation of Dominant Problems to Stages of Growth in Technology-Based New Ventures. *Academy of Management Journal*, *31*(2), 257–279. https://doi.org/10.5465/256548

Kazanjian, R. K., & Drazin, R. (1990). A stage-contingent model of design and growth for technology-based new ventures. *Journal of Business Venturing*, *5*(3), 137–150. https://doi.org/10.1016/0883-9026(90)90028-R

Kiviluoto, N. (2013). Growth as evidence of firm success: Myth or reality? *Entrepreneurship and Regional Development*, *25*(7–8), 569–586. https://doi.org/10.1080/08985626.2013.814716

Kolokas, D., Vanacker, T., Veredas, D., & Zahra, S. A. (2020). Venture Capital, Credit, and FinTech Start-Up Formation: A Cross-Country Study [in press]. *Entrepreneurship: Theory and Practice*. https://doi.org/10.1177/1042258720972652

Kumbhat, A., & Sushil. (2018). Development Stages and Scaling Issues of Startups. In S. Dhir & Sushil (Eds.), *Flexible Strategies in VUCA Markets* (pp. 3–15). Springer. https://doi.org/10.1007/978-981-10-8926-8_1

Laitinen, E. K. (2002). A dynamic performance measurement system: Evidence from small Finnish technology companies. *Scandinavian Journal of Management*, *18*(1), 65–99. https://doi.org/10.1016/S0956-5221(00)00021-X

Lee, C., Lee, K., & Pennings, J. M. (2001). Internal capabilities, external networks, and performance: A study on technology-based ventures. *Strategic Management Journal*, *22*(6–7), 615–640. https://doi.org/10.1002/smj.181

Lee, S. (2022). The myth of the flat start-up: Reconsidering the organizational structure of start-ups. *Strategic Management Journal*, *43*(1), 58–92. https://doi.org/10.1002/smj.3333

Leitch, C., Hill, F., & Neergaard, H. (2010). Entrepreneurial and business growth and the quest for a "comprehensive theory": Tilting at windmills? *Entrepreneurship: Theory and Practice*, *34*(2), 249–260. https://doi.org/10.1111/j.1540-6520.2010.00374.x

Li, H., & Atuahene-Gima, K. (2001). Product innovation strategy and the performance of new technology ventures in China. *Academy of*

Management Journal, *44*(6), 1123–1134. https://doi.org/10.2307/3069392

Li, H. (2020). New Ventures. *Oxford Bibliographies*.

Luger, M. I., & Koo, J. (2005). Defining and tracking business start-ups. *Small Business Economics*, *24*(1), 17–28. https://doi.org/10.1007/s11187-005-8598-1

Maidique, M. A., & Zirger, B. J. (1985). The new product learning cycle. *Research Policy*, *14*(6), 299–313. https://doi.org/10.1016/0048-7333(85)90001-0

Maltz, A. C., Shenhar, A. J., & Reilly, R. R. (2003). Beyond the balanced scorecard: Refining the search for organizational success measures. *Long Range Planning*, *36*(2), 187–204. https://doi.org/10.1016/S0024-6301(02)00165-6

Markman, G. D., & Gartner, W. B. (2002). Is Extraordinary Growth Profitable? A Study of Inc. 500 High–Growth Companies. *Entrepreneurship Theory and Practice*, *27*(1), 65–75. https://doi.org/10.1111/1540-8520.t01-2-00004

Marmer, M., Herrmann, B. L., Dogrultan, E., Berman, R., Eesley, C. E., & Blank, S. (2011). *Startup genome report extra: Premature Scaling* (tech. rep. No. 10). Startup Genome.

Maula, M., Autio, E., & Murray, G. (2005). Corporate venture capitalists and independent venture capitalists: What do they know, who do they know and should entrepreneurs care? *Venture Capital*, *7*(1), 3–21. https://doi.org/10.1080/1369106042000316332

Nambisan, S., & Baron, R. A. (2013). Entrepreneurship in innovation ecosystems: Entrepreneurs' self-regulatory processes and their implications for new venture success. *Entrepreneurship: Theory and Practice*, *37*(5), 1071–1097. https://doi.org/10.1111/j.1540-6520.2012.00519.x

Nason, R. S., & Wiklund, J. (2018). An Assessment of Resource-Based Theorizing on Firm Growth and Suggestions for the Future. *Journal of Management*, *44*(1), 32–60. https://doi.org/10.1177/0149206315610635

Neely, A., Gregory, M., & Platts, K. (1995). Performance measurement system design: A literature review and research agenda. *International Journal of Operations and Production Management*, *15*(4), 80–116. https://doi.org/10.1108/01443579510083622

Nguyen-Duc, A., Münch, J., Prikladnicki, R., Wang, X., & Abrahamsson, P. (2020). *Fundamentals of Software Startups: Essential Engineering and Business Aspects*. Springer. https://doi.org/10.1007/978-3-030-35983-6

Nicholls-Nixon, C. L. (2005). Rapid growth and high performance: The entrepreneur's "impossible dream?". *Academy of Management Executive*, *19*(1), 77–89. https://doi.org/10.5465/ame.2005.15841955

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd). SAGE Publications.

Pe'er, A., Vertinsky, I., & Keil, T. (2016). Growth and survival: The moderating effects of local agglomeration and local market structure. *Strategic Management Journal*, *37*(3), 541–564. https://doi.org/10.1002/smj.2331

Peña, I. (2002). Intellectual capital and business start-up success. *Journal of Intellectual Capital*, *3*(2), 180–198. https://doi.org/10.1108/14691930210424761

Pereira, V., Corradini, C., Temouri, Y., & Mellahi, K. (2020). Investigating Institutional, Economic and Social Determinants of European Regions for Firm Growth Through Employment Generation. *British Journal of Management*, *31*(1), 162–183. https://doi.org/10.1111/1467-8551.12383

Phillips, B. D., & Kirchhoff, B. A. (1989). Formation, growth and survival; Small firm dynamics in the U.S. Economy. *Small Business Economics*, *1*(1), 65–74. https://doi.org/10.1007/BF00389917

Podoynitsyna, K., Song, M., van der Bij, H., & Weggeman, M. (2013). Improving new technology venture performance under direct and indirect network externality conditions. *Journal of Business Venturing*, *28*(2), 195–210. https://doi.org/10.1016/j.jbusvent.2012.04.004

Podsakoff, P. M., Mackenzie, S. B., Bachrach, D. G., & Podsakoff, N. P. (2005). The influence of management journals in the 1980s and 1990s. *Strategic Management Journal*, *26*(5), 473–488. https://doi.org/10.1002/smj.454

Richard, O. J., Devinney, T. M., Yip, G. S., & Johnson, G. (2009). Measuring organizational performance: Towards methodological best practice. *Journal of Management*, *35*(3), 718–804. https://doi.org/10.1177/0149206308330560

Ries, E. (2011). *The Lean Startup: How today's entrepreneurs use continuous innovation to create radically successful businesses* (1st). Crown Business.

Ripsas, S., Schaper, B., & Tröger, S. (2018). *A Startup Cockpit for the Proof-of-Concept*. Springer Gabler. https://doi.org/10.1007/978-3-658-04994-2_21

Roche, M. P., Conti, A., & Rothaermel, F. T. (2020). Different founders, different venture outcomes: A comparative analysis of academic and non-academic startups. *Research Policy*, *49*(10), 104062. https://doi.org/10.1016/j.respol.2020.104062

Roure, J. B., & Maidique, M. A. (1986). Linking prefunding factors and high-technology venture success: An exploratory study. *Journal of Business Venturing*, *1*(3), 295–306. https://doi.org/10.1016/0883-9026(86)90006-6

Ruiz-Jiménez, J. M. and Ruiz-Arroyo, M. and del Mar Fuentes-Fuentes, M. (2021). The impact of effectuation, causation, and resources on new venture performance: novice versus expert entrepreneurs. *Small Business Economics*, *57*(4), 1761–1781. https://doi.org/10.1007/s11187-020-00371-7

Santisteban, J., Mauricio, D., & Cachay, O. (2021). Critical success factors for technology-based startups. *International Journal of Entrepreneurship and Small Business*, *42*(4), 397–421. https://doi.org/10.1504/IJESB.2021.114266

Santos, S. C., & Cardon, M. S. (2019). What's Love Got to Do With It? Team Entrepreneurial Passion and Performance in New Venture Teams. *Entrepreneurship: Theory and Practice*, *43*(3), 475–504. https://doi.org/10.1177/1042258718812185

Sauermann, H. (2018). Fire in the belly? Employee motives and innovative performance in start-ups versus established firms. *Strategic Entrepreneurship Journal*, *12*(4), 423–454. https://doi.org/10.1002/sej.1267

Schlichte, F., Junge, S., & Mammen, J. (2019). Being at the right place at the right time: does the timing within technology waves determine new venture success? *Journal of Business Economics*, *89*(8–9), 995–1021. https://doi.org/10.1007/s11573-019-00947-0

Shane, S. (2009). Why encouraging more people to become entrepreneurs is bad public policy. *Small Business Economics*, *33*(2), 141–149. https://doi.org/10.1007/s11187-009-9215-5

Shane, S., & Stuart, T. (2002). Organizational endowments and the performance of university start-ups. *Management Science*, *48*(1), 154–170. https://doi.org/10.1287/mnsc.48.1.154.14280

Siegel, D. S., & Wessner, C. (2012). Universities and the success of entrepreneurial ventures: Evidence from the small business innovation research program. *Journal of Technology Transfer*, *37*(4), 404–415. https://doi.org/10.1007/s10961-010-9186-3

Singh, K. (2015). *Quantitative Social Research Methods*. SAGE Publications. https://doi.org/10.4135/9789351507741

Söderblom, A., Samuelsson, M., Wiklund, J., & Sandberg, R. (2015). Inside the black box of outcome additionality: Effects of early-stage government subsidies on resource accumulation and new venture performance. *Research Policy*, *44*(8), 1501–1512. https://doi.org/10.1016/j.respol.2015.05.009

Song, L. Z., Song, M., & Parry, M. E. (2010). Perspective: Economic conditions, entrepreneurship, first-product development, and new venture success. *Journal of Product Innovation Management*, *27*(1), 130–135. https://doi.org/10.1111/j.1540-5885.2009.00704.x

Song, M., Podoynitsyna, K., van der Bij, H., & Halman, J. I. M. (2008). Success factors in new ventures: A meta-analysis. *Journal of Product Innovation Management*, *25*(1), 7–27. https://doi.org/10.1111/j.1540-5885.2007.00280.x

Sorenson, O., Dahl, M. S., Canales, R., & Burton, M. D. (2021). Do startup employees earn more in the long run? *Organization Science*, *32*(3), 587–604. https://doi.org/10.1287/orsc.2020.1371

Soto-Simeone, A., Sirén, C., & Antretter, T. (2020). New Venture Survival: A Review and Extension. *International Journal of Management Reviews*, *22*(4), 378–407. https://doi.org/10.1111/ijmr.12229

Spender, J., Corvello, V., Grimaldi, M., & Rippa, P. (2017). Startups and open innovation: a review of the literature. *European Journal of Innovation Management*, *20*(1), 4–30. https://doi.org/10.1108/EJIM-12-2015-0131

Spiegel, O., Abbassi, P., Zylka, M. P., Schlagwein, D., Fischbach, K., & Schoder, D. (2016). Business model development, founders' social capital and the success of early stage internet start-ups: a mixed-method study. *Information Systems Journal*, *26*(5), 421–449. https://doi.org/10.1111/isj.12073

Stam, W., & Elfring, T. (2008). Entrepreneurial orientation and new venture performance: The moderating role of intra- and extra-industry social capital. *Academy of Management Journal*, *51*(1), 97–111. https://doi.org/10.5465/AMJ.2008.30744031

Stuart, R., & Abetti, P. A. (1987). Start-up ventures: Towards the prediction of initial success. *Journal of Business Venturing*, *2*(3), 215–230. https://doi.org/10.1016/0883-9026(87)90010-3

Stucki, T. (2014). Success of start-up firms: The role of financial constraints. *Industrial and Corporate Change*, *23*(1), 25–64. https://doi.org/10.1093/icc/dtt008

Su, Z., Xie, E., & Wang, D. (2015). Entrepreneurial orientation, managerial networking, and new venture performance in China. *Journal of Small Business Management*, *53*(1), 228–248. https://doi.org/10.1111/jsbm.12069

Tech, R. P. G. (2018). *Financing High-Tech Startups*. Springer. https://doi.org/10.1007/978-3-319-66155-1

Ter Wal, A. L. J., Alexy, O., Block, J., & Sandner, P. G. (2016). The Best of Both Worlds: The Benefits of Open-specialized and Closed-diverse Syndication Networks for New Ventures' Success. *Administrative Science Quarterly*, *61*(3), 393–432. https://doi.org/10.1177/0001839216637849

Thiel, P., & Masters, B. (2014). *Zero to One: Notes on Startups, Or How to Build the Future*. Virgin Books.

Trailer, J. W., Hill, R. C., & Murphy, G. B. (1996). Measuring performance in entrepreneurship research. *Journal of Business Research*, *36*(1), 15–23. https://doi.org/10.1016/0148-2963(95)00159-X

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review. *British Journal of Management*, *14*(3), 207–222. https://doi.org/10.1111/1467-8551.00375

Tripathi, N., Seppänen, P., Boominathan, G., Oivo, M., & Liukkunen, K. (2019). Insights into startup ecosystems through exploration of multi-vocal literature. *Information and Software Technology*, *105*, 56–77. https://doi.org/10.1016/j.infsof.2018.08.005

Vissa, B., & Chacar, A. S. (2009). Leveraging ties: The contingent value of entrepreneurial teams' external advice networks on Indian software venture performance. *Strategic Management Journal*, *30*(11), 1179–1191. https://doi.org/10.1002/smj.785

Wall, T. D., Michie, J., Patterson, M., Wood, S. J., Sheehan, M., Clegg, C. W., & West, M. (2004). On the validity of subjective measures of company performance. *Personnel Psychology*, *57*(1), 95–118. https://doi.org/10.1111/j.1744-6570.2004.tb02485.x

Wang, D., Pahnke, E. C., & McDonald, R. M. (2022). The past is prologue? Venture-capital syndicates' collaborative experience and start-up exits. *Academy of Management Journal*, *65*(2), 371–402. https://doi.org/10.5465/amj.2019.1312

Wang, M.-C., & Chen, M.-H. (2016). The more, the better? The impact of closure collaboration network and network structures on technology-based new ventures' performance. *R and D Management*, *46*, 174–192. https://doi.org/10.1111/radm.12129

Wasserman, N. (2017). The throne vs. the kingdom: Founder control and value creation in startups. *Strategic Management Journal*, *38*(2), 255–277. https://doi.org/10.1002/smj.2478

Wennberg, K., Wiklund, J., DeTienne, D. R., & Cardon, M. S. (2010). Reconceptualizing entrepreneurial exit: Divergent exit routes and their drivers. *Journal of Business Venturing*, *25*(4), 361–375. https://doi.org/10.1016/j.jbusvent.2009.01.001

Wiklund, J., & Shepherd, D. (2005). Entrepreneurial orientation and small business performance: A configurational approach. *Journal of Business Venturing*, *20*(1), 71–91. https://doi.org/10.1016/j.jbusvent.2004.01.001

Witt, P. (2004). Entrepreneurs' networks and the success of start-ups. *Entrepreneurship and Regional Development*, *16*(5), 391–412. https://doi.org/10.1080/0898562042000188423

Woolley, J. L., & MacGregor, N. (2021). The Influence of Incubator and Accelerator Participation on Nanotechnology Venture Success. *Entrepreneurship: Theory and Practice*. https://doi.org/10.1177/10422587211024510

Yua, S. (2020). How do accelerators impact the performance of high-technology ventures? *Management Science*, *66*(2), 530–552. https://doi.org/10.1287/mnsc.2018.3256

Zahra, S. A. (2021). The Resource-Based View, Resourcefulness, and Resource Management in Startup Firms: A Proposed Research Agenda. *Journal of Management*, *47*(7), 1841–1860. https://doi.org/10.1177/01492063211018505

Zahra, S. A., Bogner, W. C., Bantel, K., Hill, R. C., Nielsen, A. P., & Zahra, P. H. (2000). Technology strategy and software new ventures' performance: Exploring the moderating effect of the competitive environment. *Journal of Business Venturing*, *15*(2), 135–173. https://doi.org/10.1016/S0883-9026(98)00009-3

Zhang, J., & Gu, Q. (2021). Turning a curse into a blessing: Contingent effects of geographic distance on startup–VC partnership performance. *Journal of Business Venturing*, *36*(4), 106108. https://doi.org/10.1016/j.jbusvent.2021.106108

Junior Management Science

www.jums.academy
ISSN: 2942-1861

Mit Selbstoptimierung zum individuellen Glück?

Self-Optimization for Individual Happiness? Title of the research article

Lea Krähenmann

University of St.Gallen

Abstract

In our secular society, individuals are urged to find meaning in their lives despite limited opportunities. While inner hero's journeys convey a sense of purpose, depth, and belonging, apparent promises of happiness, such as self-optimization, are gaining increasing traction. This exploratory study aims to investigate whether students perceive self-optimization as a path to a good life and how it relates to transformative journeys. Utilizing a combination of semi-structured interviews and a quantitatively analyzed questionnaire, the societal views of students, their practices and motivations for self-optimization, notions of a good life, and interconnections are explored. In the post-heroic, secular, and performance-oriented society described by students, the majority expects to self-improve. The analysis reveals that only a small fraction experiences transformative journeys, while others strive for happiness in seemingly transformative developments like self-optimization or other life projects. The hypothesis that this pursuit of happiness is based on incomplete understandings of meaning and self-awareness should be examined in further studies.

Zusammenfassung

In unserer säkularen Gesellschaft sind Menschen trotz begrenzter Möglichkeiten aufgefordert, ihrem Leben Sinn zu geben. Während innere Heldenreisen ein Gefühl von Sinn, Tiefe und Zugehörigkeit vermitteln, gewinnen scheinbare Glücksversprechen wie Selbstoptimierung zunehmend an Bedeutung. Diese explorative Studie zielt darauf ab zu untersuchen, ob Studierende Selbstoptimierung als Weg zum erfüllten Leben betrachten und wie sich dies zu transformativen Entwicklungsgeschichten verhält. Mithilfe von semistrukturierten Interviews und einem quantitativ ausgewerteten Fragebogen werden das Gesellschaftsbild der Studierenden, ihre Selbstoptimierungspraktiken und -motive, Vorstellungen eines gelungenen Lebens sowie Zusammenhänge erforscht. In der postheroischen, säkularen und leistungsorientierten Gesellschaft erwartet die Mehrheit der befragten Studierenden von sich selbst, an sich zu arbeiten. Die Auswertung zeigt, dass nur wenige Befragte transformative Entwicklungsgeschichten erleben, während andere in vermeintlich transformativen Entwicklungen wie Selbstoptimierung oder anderen Lebensprojekten nach Glück streben. Die These, dass dieses strebende Glück auf unvollständigem Sinn- und Selbstkenntnisverständnis beruht, sollte in weiterführenden Studien überprüft werden.

Keywords: finding meaning; good life; self-improvement; self-optimization; transformative hero journey

Es sei den Befragten für ihre Teilnahme an dieser Bachelorarbeit gedankt. Nebst dem zeitlichen Aufwand, den die Befragten aufgrund der Befragung auf sich nahmen, wird insbesondere das Vertrauen der Befragten der Interviewerin gegenüber sehr geschätzt. Für die umfassende Unterstützung während sowie der Begutachtung dieser Bachelorarbeit sei Carole Inäbnit, Sven Glinz, Annette Arbitter und Beat Krähenmann gedankt. Der Dank gilt ebenso dem Betreuer dieser Arbeit, Herrn Professor Kolmar, welcher mit konstruktiver Kritik und fachlichen Ratschlägen in spannenden Gesprächen den Prozess dieser Arbeit unterstütze und bereicherte.

DOI: https://doi.org/10.5282/jums/v9i1pp1341-1383

1. Einleitung

”Ich glaube nicht an Gott, aber ich vermisse Ihn.” (Barnes, 2008). Das Zitat von Julian Barnes ist charakteristisch für Menschen, die in einer säkularen Gesellschaft leben. Eine säkulare Gesellschaft ist insofern schizophren, als sie einerseits Bürger*innen fern von Religion und anderen transzendenten Erfahrungen hält und andererseits erwartet, dass diese individuell, basierend auf ihrem eigenen Wissen und ihrer Leistung, ihrem Leben Bedeutung geben (C. Taylor, 2007, S. 727). Der epistemische Rahmen der säkularen Gesellschaft bietet jedoch keinen Raum für die Befriedigung von menschlichen Bedürfnissen nach Zugehörigkeit, Sinn und Tiefe - wodurch diese jedoch nicht verschwinden (Kolmar, 2021, S. 233).

Im Buch ”der Heros in tausend Gestalten” zeigt der Mythenforscher Campbell (1953), dass menschliche Geschichten über Zeit und Kulturen einer verallgemeinerbaren Struktur der sogenannten Heldenreise folgen. Die Heldenreise steht symbolisch für eine innere, psychologische Entwicklungsgeschichte, die Menschen durch die daraus gewonnene Selbst- und Menschenkenntnis Sinn erfahren lässt. Campbell’s Analyse der heutigen (westlichen) Gesellschaft ergibt allerdings, dass der Sinn nicht länger im Kollektiv, gar der Welt, sondern im Individuum zu finden ist. Da sich das Individuum des Sinnes unbewusst ist, besteht ”das Problem der modernen Welt darin, ihr einen Sinn zu verleihen” (Campbell, 1953, S. 371).

Spiritualität, Transzendenz und Gemeinschaft sind laut der Positiven Psychologie wichtige Wege, Sinn zu erfahren und tragen erheblich zur subjektiven Lebenszufriedenheit bei (Diener & Seligman, 2002, 2004; Seligman, 2002). Gleichzeitig halten sich seit dem antiken Griechenland zu Glück führende Leitideen der Verbesserung von sich selbst und der Welt (Balandis & Straub, 2018). Die Idee hat sich, wie einige Soziologen und Soziologinnen argumentieren, vielmehr zu einem Zwang des Selbstoptimierens und -investierens gewandelt (Balandis & Straub, 2018; Mühlhausen & Wippermann, 2013). Im proklamierten Zeitalter der Selbstoptimierung wird dem Individuum in vielen Praktiken wie beispielsweise dem Konsum von Ratgebern, leistungssteigernder Substanzen oder der Durchführung von Schönheitsoperationen, Glück versprochen (Röcke, 2021, S. 7). Selbstoptimierende Praktiken enthalten auch destruktives Potential. Nicht alle Individuen sind in der Lage, sich erfolgreich den gesellschaftlichen Anforderungen der Selbstoptimierung anzunehmen und das ”unternehmerische Selbst” (Bröckling, 2013) droht zum ”erschöpften Selbst” (Ehrenberg, 2015) zu werden. Eine Zunahme von Stress, Depressionen und anderen psychischen Erkrankungen ist die Folge.

Junge Menschen haben in den neusten Schweizer Studien zur psychischen Gesundheit besonders schlecht abgeschnitten (Barrense-Dias et al., 2021). Studierende wiesen in einer vorangehenden Studie durchschnittlich schlechtere mentale Gesundheitswerte auf als die Gesamtbevölkerung der gleichen Alterskategorie (Bundesamt für Statistik, 2018, S. 12). Erste Studien untersuchen die Lebenszufriedenheit von Studierenden (Jakubasch, 2017; Schramek, 2012; Zeidner, 2021), deren Erfolgsnarrative (Donaldson et al., 2000; Picton et al., 2018) sowie den Einfluss ihrer Selbstverständnisse auf den Lebenssinn (Venter, 2016). Forschungslücken bestehen in der Wahrnehmung von Selbstoptimierungszwängen Schweizer Studierenden, der daraus abgeleiteten Selbstoptimierungspraktiken und der subjektiven Einordnung der Selbstoptimierung in Vorstellungen eines gelingenden Lebens.

Ziel dieser explorativen Arbeit ist es, aufzudecken, wie und aus welcher Überzeugung Studierende der Wirtschaftswissenschaften Selbstoptimierung betreiben, welche Bedeutung sie dieser zur Erreichung eines gelingenden Lebens beimessen und inwiefern ihre individuellen Selbstoptimierungsnarrative als Ersatz für fehlende grössere Sinnzusammenhänge in der Gesellschaft dienen. Die Beantwortung der Forschungsfrage:

> *Inwiefern konkurrieren Selbstoptimierungsnarrative und -praktiken von Studierenden der Wirtschaftswissenschaften mit Heldenreisen, die dem gelingenden Leben zuträglich sind?*

ist insbesondere für Studierende und gesellschaftliche Institutionen wichtig, da sich diese gegebenenfalls von der Selbstoptimierung eine transformative Reise zum gelingenden Leben erhoffen und aufgrund dessen keine tiefergreifende innere Heldenreise in Angriff nehmen. Zur Beantwortung der Forschungsfrage werden in einem ersten Schritt theoretisch Überschneidungen und Unterschiede von Selbstoptimierung und Heldenreisen in der säkularen Gesellschaft erarbeitet. Mithilfe einer Kombination quantitativer und qualitativer Methoden wurden die Selbstoptimierungspraktiken und -narrative der Studierenden sowie deren Einordnung in die Vorstellung eines gelingenden Lebens erfragt. Auf die Erläuterungen zur Methodik folgen die Illustration und Diskussion der Resultate. Diese explorative Arbeit schliesst mit Limitationen und Vorschlägen zu weiterführenden Studien.

2. Theoretische Grundlagen

2.1. Sinn in einer säkularen Gesellschaft

C. Taylor (2007) untersucht in seinem Buch ‘A secular age’, den Wandel vom christlichen Einheitsglauben in europäischen beziehungsweise westlichen[1] Ländern zur Auflösung dessen in kleinere und fehlende Glaubenseinheiten. Eine säkulare Welt ist somit nicht gleichzusetzen mit einer ungläubigen Welt, sondern mit einer in welcher, aufgrund des vorherrschenden wissenschaftlichen Bezugsrahmens,

[1] Westlich wird in dieser Arbeit verwendet um die kultur-historischen Veränderungen der westlich gelegenen Kontinente wie Europa und Nordamerika zu beschreiben. Dazu gehört insbesondere die Säkularisierung, die unter anderem durch eine Trennung von Kirche und Staat sowie religiöser Vielfalt und Abnahme der Religiosität im Allgemeinen charakterisiert ist.

eine Fragmentalisierung und Pluralisierung des Glaubens stattgefunden hat (C. Taylor, 2007, S. 727). Religion muss sich heutzutage im wissenschaftlichen Kontext einordnen und nicht mehr umgekehrt (Kolmar, 2021, S. 47).

Zudem ist der heutige epistemische Rahmen von Entzauberung[2] geprägt. J. Smith (2019) drückt Taylor's Entzauberung als Bedeutungsverlust von Transzendenz und Unendlichkeit aus. Es entsteht ein sogenannter exklusiver Humanismus, "eine Art, in der Welt zu sein, die die tiefsten Quellen des Sinns nur in Bezug auf das menschliche Leben und nicht in Bezug auf eine Realität ausserhalb oder jenseits des menschlichen Lebens verortet" (Hondo, 2013). Wie in der Einleitung beschrieben wird gesellschaftlich vom Individuum erwartet, dem eigenen Leben durch eigene Leistung und eigenes Wissen einen Sinn zu verleihen. Durch die Etablierung eines wissenschaftlichen Gesellschaftskontextes wurden nach dem Sinn fragende 'wozu' durch 'wie' Fragen verdrängt (Kolmar, 2021, S. 48).

Die Menschen suchen im naturalistischen Kontext vergeblich nach Antworten auf 'wozu' Fragen, denn deren Nichtbeantwortung unterdrückt nicht deren Existenz. Der amerikanische Psychologe Maslow bestätigte das menschliche Bedürfnis des Sinnverspüren des eigenen Lebens, indem er kurz vor seinem Tod Selbsttranszendenz als oberstes menschliches Bedürfnis hinzufügte. Selbsttranszendenz beschreibt ein Gefühl, sich selbst als sinnvollen Teil etwas Grösserens verstehen zu können (Venter, 2016, S. 3). Der immanente Rahmen verunmöglicht Transzendenz, indem ein gelingendes Leben normativ auf Selbstverantwortung, Autonomie und Selbstdisziplin basiert (Kolmar, 2021, S. 50). Menschen suchen in säkularen Gesellschaften entsprechend in Lebensprojekten wie Familiengründung, Karriere oder dem materiellen Konsum nach einem erfüllenden Leben, was sich letztendlich als Illusion erweist (Kolmar, 2021, S. 49; J. Smith, 2019; C. Taylor, 2007, S. 14). Der Psychoanalytiker Jacques Lacan spricht davon, dass das Individuum durch die notwendige Anpassung an soziale Konventionen[3] ein Teil des eigenen Wesens verliert. In seiner pessimistischen Deutung der Zukunft leben alle Menschen aufgrund dieser unerreichbaren Vollständigkeit[4] mit einer "nicht zu füllenden Leere im Zentrum ihrer Existenz" (Kolmar, 2021, S. 178). Es ist diese Leere, die laut Lacan, zur Triebkraft kapitalistischer Gesellschaften wird: "Konsum und Fortschritt werden zu Versuchen, die durch sie letztendlich unerfüllbaren Wünsche nach Sinn zu befriedigen" (Kolmar, 2021, S. 48). Das Individuum geht in der Erforschung neuer Lösungen verloren, weil das Ziel unerreichbar ist, was eine Voraussetzung für den sogenannten Nova-Effekt darstellt (C. Taylor, 2007, S. 423).

Dass das eigene Leben nicht mit zureichend Bedeutung, Sinn und Zugehörigkeitsgefühl versehen werden kann, führe "zu einem Gefühl der Leere, einem Mangel im Zentrum der Existenz, einem Verlust an Gründen und damit einem Verlust eines Grundes, auf dem man stehen könnte" (Kolmar, 2021, S. 48). Neben der persönlichen Enttäuschung über die Nichterfüllung des Wunsches, seinem eigenen Leben einen Sinn zuzuschreiben, wird das gesellschaftlich als Leistungsversagen des Individuums geächtet. Der gesellschaftliche und psychologische Druck führt letztendlich zu einer mehr oder weniger problematischen Entladung der angestauten Energie, dem Nova-Effekt. Ein weniger problematisches Beispiel stellt das Einnehmen einer anderen Rolle am Karneval dar. Problematisch sieht Taylor die Radikalisierung und Führung höchst gewalttätiger 'heiliger' Kriege, die nicht Gottes wegen sondern des gemeinschaftlich wahrgenommenen Lebenssinns wegen geführt werden (C. Taylor, 2007, S. 300, 625).

Im Rahmen dieser Arbeit wird geprüft, ob Studierende fälschlicherweise in der Selbstoptimierung den gesellschaftlich fehlenden Sinn suchen. Die nächsten Kapitel widmen sich entsprechend dem Phänomen der Selbstoptimierung.

2.2. Selbstoptimierung

Ziel dieses Kapitels ist es, den Begriff Selbstoptimierung für die Arbeit einzugrenzen. Dafür wird in einem ersten Schritt der relativ neue Begriff angenähert sowie in einem zweiten positive und negative Aspekte ausgeführt. Schliesslich wird die Verwendung von Selbstoptimierung in dieser Arbeit dargelegt.

2.2.1. Begriffsannäherung

Im Jahre 2014 hatten 61% der deutschen Bevölkerung von Selbstoptimierung weder gehört noch gelesen ("Das optimierte Selbst", 2014). Trendforscher*innen sprechen derweil vom 21. Jahrhundert als dem der Selbstoptimierung und Soziologen*innen sprechen von Optimierungsgesellschaften (Balandis & Straub, 2018; Mühlhausen & Wippermann, 2013).

Die Verbesserbarkeit oder gar Perfektionierbarkeit stellt den Anfang jedes Optimierens dar: Prinzipiell können Dinge oder Personen zum Besseren verändert werden. Der Begriff des Optimierens geht zurück auf das lateinische 'optimus', was als der Beste oder der Tüchtigste übersetzt wird. Optimierendes Verhalten sollte demnach zum bestmöglichen Zustand, dem sogenannten 'Optimum', führen. Das 'Selbst' stellt allgemein einen Bezug auf die eigene Person dar. Das Selbst in Selbstoptimierung beschreibt ein Ichbewusstsein,

[2] Der Begriff Entzauberung wurde von Max Weber geprägt. Er beschreibt damit eine zunehmende Rationalisierung, und Intellektualisierung der Welt, welche auf Kosten mystischer Weltanschauungen Wissenschaftliche ermöglicht. Für weitere Ausführungen des Konzepts sei hier aufgrund des begrenzten Rahmens auf (Weber, 2002) verwiesen.

[3] Dazu gehören beispielsweise soziale Interaktionen, die Sprache sowie die Kultur. Lacan gibt diesen den Namen 'das Symbolische'.

[4] Lacan beschreibt dies mit folgenden Begriffen: 'Das Imaginäre' umfasst Selbstbilder sowie Beziehungen zu anderen Personen. Es beinhaltet somit Aspekte wie Identität, Fantasien, Begehren und das Verhältnis zum eigenen Körper. 'Das Reale' stellt laut Lacan die Ebene der Wirklichkeit dar, die sich jenseits von Symbolen, Sprache und Konstruktionen befindet (Lacan, 1977, 1992). Imaginäres und Reales werden sich niemals übereinkommen. 'Das Ding' ist Ausdruck der Hoffnung, "die eigene Leere füllen zu können" (Kolmar, 2021, S. 177). Es stellt jedoch ein "verlorenes Objekt dar, welches nicht angeeignet werden kann, nie eigen war. Es ist der Platzhalter für die Leere im Symbolischen" (Kolmar, 2021, S. 178).

welches für die Entscheidung, intentional an sich zu arbeiten, als Voraussetzung gilt. Durch den Selbstbezug der Selbstoptimierung wird das Individuum Subjekt und Objekt zugleich (Röcke, 2021, S. 52). Aus soziologischer Perspektive ergibt sich das Selbst aus komplexen Wechselbezügen zwischen eigenen Handlungen einerseits und dem sozialen Umfeld andererseits und kann deshalb nicht einheitlich sein (Röcke, 2021, S. 182).

Die Selbstoptimierung unterscheidet sich durch ihre zeitliche Unbeschränktheit und den freiwilligen instrumentellen Selbstbezug von der Selbstverbesserung: Es wird verbessert, bis nichts mehr verbessert werden kann - sodass ein Selbstverbesserungsende wieder zum Anfang wird. Oftmals verschieben sich die Ziele bereits während den fortschreitenden Handlungen. "Die Vorläufigkeit des relativ Besten trennt die Optimierung von der Perfektionierung, also vom Perfekten oder Vollkommenen" (Girkinger, 2019, S. 16). Laut Balandis und Straub (2018) hat die "Optimierung kein endgültiges Ende, nichts Definitives vor Augen" (C. Taylor, 2007, S. 300, 625). Die vorausgesetzte Unabschliessbarkeit hebt die Prozessdimension der Selbstoptimierung trotz (Schein-)Zielen in den Vordergrund (Balandis & Straub, 2018, S. 4).

Duttweiler, deren Fokus auf der alltäglichen Selbstoptimierung liegt, beschreibt Selbstoptimierung als "kleine Modifikationen der alltäglichen Lebensführung hin zu einem glücklicheren, fitteren oder gesünderen Leben" und nicht als "radikale Verwandlung zu einem Neuen oder einem perfekten Menschen". Das Leben wird zur "ewigen Baustelle, denn immer wieder werden neue Ziele anvisiert und ständig 'Ausbesserungen' in diversen Bereichen vorgenommen" (Duttweiler, 2016, S. 27). Ein weiteres Merkmal der Selbstoptimierung stellt das Handeln im [angenommenen] eigenen Interesse dar (Röcke, 2021, S. 25, 168).

Die Selbstoptimierung ist inhaltlich nicht definiert. Selbstoptimierung hat weder einen intrinsischen Zweck noch eineRegel zur Messung des Fortschritts. Wie Röcke schreibt kann die "leere Hülle (der Selbstoptimierung) beliebig gefüllt werden" (Röcke, 2021, S. 178). Zahlreiche Dimensionen des Selbst können optimiert werden. Dazu zählen beispielweise physische, gesundheitliche, soziale, psychische und geistige Zustände, Eigenschaften sowie Kompetenzen, Charakterzüge, Fähigkeiten und Prozesse. Inhaltlich kann die Selbstoptimierung entsprechend individuell variieren (Heßdörfer, 2018). Gemeinsam bleibt den Selbstoptimierungsstreben die Intention das, gegeben den Möglichkeiten, „situativ bestmögliche, über dem 'normalen Mass' liegende Ergebnis zu erreichen, das aber immer wieder 'getoppt' werden kann. Die grundlegende Intention ist es, besser zu werden, aber nicht im Sinne eines 'besseren Scheiterns', sondern des 'besser Bessermachens'" (Röcke, 2021, S. 9). Neuste Befragungen zeigen, dass sich die Ziele der Selbstoptimierer*innen von leistungskonnotierten Begriffen wie "höher, schneller, weiter" zu innenorientieren Wertehaltungen und Bereichen wie Freiheit, Gesundheit sowie Glück verschiebt (Mühlhausen, 2016b, S. 4). Auch Modifikationen am Körper können "dem Ausdruck der eigenen Identität und Persönlichkeit dienen" und ein Aspekt von Selbstoptimierung sein. Dieser objekthafte Erlebensaspekt des Körpers "sei bei diesen Personen deutlich stärker ausgeprägt als bei Personen ohne Körpermodifikationen " (Röcke, 2021, S. 186).

Im Duden wird die Verwendung des Wortes Selbstoptimierung als "leicht abwertend" beschrieben. Selbstoptimierung wird als "jemandes [übermässige] freiwillige Anpassung an äussere Zwänge, gesellschaftliche Erwartungen oder Ideale" definiert (Duden, 2023c). Hierbei wird klar, dass Selbstoptimierung kein neutraler Begriff darstellt. Röcke (2021, S. 201) beschreibt Selbstoptimierung als Verwobenheit zwischen "Freiheit und Zwang, sowie Autonomie und Heteronomie". Im nächsten Unterkapitel werden entsprechend positive und negative Aspekte der Selbstoptimierung beleuchtet.

2.2.2. Positive und negative Aspekte der Selbstoptimierung

An sich selbst zu arbeiten ist ein fundamentales menschliches Bedürfnis. Kleine und grössere Verbesserungen menschlicher Fähigkeiten und Aussehen haben eine lange Tradition. Die Verbindung zu Glück beleuchtete bereits Foucault im Rahmen der Technologien des Selbst[5] (1988). Als positiv wird unter anderem die grundsätzliche Offenheit der Selbstoptimierungsinhalte gewertet. Das westliche Ideal der individuellen Selbstbestimmung wird in Selbstoptimierungsbestrebungen sichtbar: Jede*r habe die Freiheit, die eigenen Werte und Ziele zu definieren (Gamm, 2013, S. 48). Selbstoptimierungspraktiken können zu Erfolgen führen und somit einen motivierenden Effekt haben. Die Erreichung eigener Ziele kann zudem Gefühle von Selbstwirksamkeit hervorrufen. Stolz, Anerkennung, eine bessere Stimmungslage sowie ein erhöhtes Selbstbewusstsein und Selbstwertgefühl sind weitere Vorteile, die selbstoptimierende Praktiken (zumindest vorübergehend) mit sich bringen können (Balandis & Straub, 2018). Laut Röcke können Selbstoptimierungspraktiken auch mit anderen positiv bewerteten Praktiken wie der Selbstheilung, Selbstsorge und Selbstformung verbunden sein (Röcke, 2021, S. 64). Die eigene Energie dafür aufzuwenden, an sich selbst zu arbeiten, kann zudem in einer dynamischen und komplexen Welt befreiend und befriedigend sein. Einerseits kann das Mitmachen bei der Selbstoptimierung den Anschluss an eine dynamische Umwelt ermöglichen und andererseits können scheinbar unerschöpflich viele positive Erfahrungen durch Selbstoptimierungspraktiken gemacht werden, während die Umwelt nicht im gleichen Masse beeinflussbar ist. Allerdings festigt umgekehrt die Selbstoptimierung der Individuen die gesellschaftlichen Prozesse erneut, denn "eine auf dynamischer Steigerung basierende gesellschaftliche Struktur setzt

[5] Unter Technologien des Selbst fasst Foucault Techniken und Praktiken zusammen, die es Individuen ermöglicht, sich selbst zu reflektieren, kontrollieren und zu einer besseren Version zu verändern. Diese Technologien umfassen beispielsweise Selbstbeobachtung, Schreibpraktiken, spirituelle Übungen und andere Formen der Selbsterforschung, die im Rahmen sozialer und kultureller Normen und Machtverhältnisse wirken (Foucault, 1988).

demnach Individuen voraus und bringt solche zugleich hervor, die ebenfalls auf permanente Aktivität, Anpassung und Steigerung setzen" (Röcke, 2021, S. 221).

Kritiker*innen der Selbstoptimierung stufen diese als vermeintlich frei ein: Vielmehr sehen sie Selbstoptimierung als aufgezwungen und selbstausbeuterisch. Selbstoptimierung führt nicht immer zum gewünschten Resultat und kann beispielsweise aufgrund der Unabgeschlossenheit ein Ausbrennen zur Folge haben oder durch andere negative Konsequenzen kontraproduktiv wirken. Abgelehnt wird die Selbstoptimierung auch, weil der Optimierungs-Begriff im ökonomischen Bereich Anwendung findet, wo er mit Effizienz- und Produktivitätssteigerung verbunden wird, was wiederum als inadäquate Beschreibungen für menschliche Lebensstile und Selbstbezüge angesehen wird. Nachfolgende Soziologen sprechen direkt von der Anwendung der kapitalistischen Logik auf Personen und Dinge. Laut Reckwitz (2017, S. 8) leben wir "nicht nur im industriellen sondern auch kulturellen Kapitalismus": Menschen müssten einen Mehrwert aus sich herausholen, was zum einen die Steigerung von Effizienz und Effektivität, zum anderen eine "unbegrenzter Kapitalakkumulation durch den Einsatz formell friedlicher Mittel" (Boltanski & Chiapello, 2003, S. 39) beinhaltet. Das Kapital stellen nun der eigene Körper und Geist dar. Zum einen können Personen beispielsweise durch das Erlangen von Bildungsabschlüssen oder der Arbeit am eigenen Körper Kapital akkumulieren. Zum anderen gilt die Kapitallogik transaktional: Kapital wird in Beziehungen ausgetauscht, um die eigenen Ziele und Bedürfnisse zu befriedigen. Der französische Soziologe Pierre Bourdieu führte das Konzept des "Sozialen Kapitals" ein.[6] Die Ökonomisierung des Sozialen und des Selbst ist unter anderem von Ulrich Bröckling im Buch "Das unternehmerische Selbst" (2013) ausführlich beschrieben worden. Im Anhang A ist ein Kapitel der Entwicklung zum Zwang der Selbstinvestition gewidmet[7].

Die Kritik am neoliberalen Wirtschaftssystem sowie an der Ökonomisierung der menschlichen Lebensweise ist nicht unbegründet: Sie können zu "Formen der Ausschliessung, gesellschaftlichen Druck und grosses psychisches Leid verursachen, die körperlich-leibliche Struktur nicht nur verändern, sondern auch dauerhaft verletzen sowie zu Formen der Entfremdung, Verdinglichung und Ausbeutung führen" (Röcke, 2021, S. 229). Die Selbstoptimierung als Symptom einer Wettbewerbsgesellschaft abzustempeln, stellt laut Foucault allerdings eine reduktionistische Sichtweise dar. "Denn Selbstoptimierungspraktiken [...] müssen nicht notwendig unter ökonomischem Druck stattfinden, sondern können auch im Rahmen einer 'ästhetik der Existenz' ein experimentell-spielerisches Selbstverhältnis zum Ausdruck bringen oder zu einer 'Selbstexpertisierung' und Emanzipation von standardisierten medizinischen oder gesellschaftlichen Normen führen" (Meissner, 2016, S. 333, 340).

Zusammenfassend kann gesagt werden, dass Selbstoptimierung als spätmodernes Phänomen als solche nicht gut oder schlecht ist. Befürworter*innen und Kritiker*innen erfassen mit ihren Sichtweisen jeweils nicht das ganze Bild. Wie beispielsweise Foucault's Werke zeigen, sind kritische öffentliche Diskussionen für den konstruktiven Umgang mit gesellschaftlichen Normen und davon ausgehender Druck zentral, diese aber nicht per se schlecht (Fenner, 2020; Foucault, 1977). Die Erkenntnisse dieser Arbeit, insbesondere, ob Studierende ihr persönliches Glück (vergeblich) in der Selbstoptimierung suchen, könnten neue Diskussionspunkte ermöglichen. Bevor mögliche Spannungsverhältnisse zwischen Selbstoptimierungsnarrativen und Heldenreisen untersucht werden können, wird im nächsten Kapitel die Verwendung von Selbstoptimierung in dieser Arbeit abgegrenzt.

2.2.3. Begriffsabgrenzung

Selbstoptimierung steht in dieser Bachelorarbeit "für eine individualistische Strategie der instrumentell ausgerichteten Selbstüberbietung [...], die prinzipiell kein Ende hat" (Röcke, 2021, S. 179-180). Eine Person arbeitet kontinuierlich aus eigenem Antrieb in mindestens einem Bereich des eigenen Lebens, zum Beispiel dem eigenen Aussehen, der eigenen Produktivität, Gesundheit, sozialen Beziehungen und Kompetenzen, der eigenen Alltagsgestaltung oder dem persönlichen Glück an sich. Die Arbeit an sich selbst muss zudem von einer Idee des Fortschritts geleitet sein sowie im Auge der Durchführenden nicht als erledigt gekennzeichnet werden können: Ein erledigtes Ziel führt in das Nächste. Das routinemässige Schneiden lassen der Haare, ohne jedes Bestreben, durch gezielte Arbeit am eigenen Körper einem bestimmten Schönheits- oder Lifestyle-Ideal näherzukommen, zählt entsprechend in dieser Arbeit nicht als Selbstoptimierung, da der Fortschrittsgedanke fehlt. Ebenso wenig zählen notwendige Veränderungen an sich wandelnde Umstände, ohne das eigene Bestreben, daraus einen Vorteil zu ziehen, zur Selbstoptimierung.

Alle Tätigkeiten, die einfach so durchgeführt werden und somit einen Selbstzweck innehaben, gehören zugleich nicht zur Selbstoptimierung. Zum Beispiel sind Praktiken, in denen eine Person Kenntnis über sich selbst erlangt, dann keine Selbstoptimierung, wenn es keine Absicht hinter deren Durchführung gibt, ausser, dass es sich in dem Moment gut

[6] Nach Bourdieu besteht das soziale Kapital aus den Ressourcen, die eine Person aufgrund ihrer sozialen Beziehungen zur Verfügung hat. Diese Ressourcen können unterschiedlicher Art sein, z.B. Informationen, Einfluss, Solidarität oder finanzielle Mittel. Der Austausch von sozialem Kapital findet in sozialen Beziehungen statt und dient dazu, das Kapital zu erhöhen oder zu erhalten. In einer Beziehung zwischen zwei Individuen oder Gruppen können ähnliche Austauschbeziehungen stattfinden. Die transaktionale Kapitallogik sieht solche Beziehungen als Austausch von Ressourcen an, die beiden Parteien zugutekommen sollen. Dabei können sowohl monetäre als auch nicht-monetäre Ressourcen ausgetauscht werden. Neben dem sozialen Kapital beschreibt Bourdieu Zeit und materielle Mittel als ökonomisches Kapital sowie Bildungs- und Handlungswissen als kulturelles Kapital. Symbolisches Kapital beschreibt die Macht, die Individuen oder Gruppen aufgrund ihrer sozialen Anerkennung ausüben können (Jurt, 2012).

[7] Im Rahmen dieser Arbeit kann der Transformation des 'an sich Arbeitens' hin zur Selbstinvestition nicht volle Rechnung getragen werden. Es sei auf die zitierte Masterarbeit von Sprenger (2017) verwiesen.

oder notwendig anfühlt.[8]

Wenn aber die Selbstkenntnis eine Person beispielsweise zu einer besseren Führungsperson macht und dann wiederum die eigenen Berufs-Chancen erhöht und aus diesem Grund die Selbstkenntnis aktiv erforscht wird, hat dies zwar einen Zweck für die eigene Person, also einen Eigenzweck, nicht aber einen Selbstzweck 'in sich'.[9] Würde die Selbstkenntnis erarbeitet werden, um die vorgesetzte Person zu beeindrucken, kann sogar von einem Fremdzweck gesprochen werden. Tätigkeiten, die das Individuum in den Augen anderer optimierter erscheinen lassen, zählen zur Selbstoptimierung. Lediglich Tätigkeiten, die einen Selbstzweck innehaben, stellen keine Selbstoptimierung dar.

Selbstoptimierung wird oftmals über konstante Messung und systematische Feedbackschlaufen gesteuert. In dieser Arbeit zählen auch Selbstoptimierungspraktiken, die einmal zu Eigen- oder Fremdzwecken verinnerlicht wurden und nicht systematisch überprüft werden oder nicht messbar überprüft werden können, als Selbstoptimierung. Selbstoptimierung muss nicht allein durchgeführt werden. Ratgeber oder persönliche Berater*innen können zur Selbstoptimierung hinzugezogen werden.

Die in diesem Überkapitel beschriebene Selbstoptimierung könnte von Studierenden als für ihr Leben zentrale Entwicklungsgeschichte wahrgenommen werden. In diesem Fall besteht die Möglichkeit, dass Selbstoptimierungsnarrative transformative Entwicklungen in Form von Heldenreisen in den Hintergrund drängen. Inwiefern dazu Tendenzen bestehen, ist Forschungsgegenstand dieser explorativen Arbeit. Vorerst werden Heldenreisen im nächsten Kapitel theoretisch erschlossen und mit Selbstoptimierugsnarrativen verglichen.

2.3. Heldenreise

2.3.1. Monomythos und die innere Heldenreise

Die Begriffe Held*in und Heldentum werden in dieser Arbeit mit Bezug auf die innere Heldenreise, wie sie unter anderem im Buch 'Der Heros in tausend Gestalten' von Joseph Campbell aufgezeigt wurde und in diesem Kapitel beschrieben wird, verwendet.[10] Der Mythenforscher und Professor untersuchte mündlich weitergegebene Geschichten und Lebenserfahrungen aller Kulturen und Zeitalter auf Gemeinsamkeiten und die daraus abgeleiteten fundamentalen menschlichen Bedürfnisse. Im Buch verglich Campbell alle verfügbaren Weisheitslehren, Märchen, Mythen, Religionen und Sagen mit besonderem Fokus darauf, was ein gelingendes Leben ausmacht. In Campbell's Worten bereiten Mythen "ein Feld, in welchem sich ein Individuum verorten kann und welche es durch die Phasen des Lebens tragen" (Campbell, 2004, S. xvi). Campbell erkannte eine übereinstimmende Erzählung; den sogenannten Monomythos[11] (Kolmar, 2021, S. 121), die durch "innere psychologische Zusammenhänge" erklärt werden kann (Campbell, 1972, S. 153).

"Overall, the hero's outer journey is a representation of an inner, psychological journey that involves leaving one condition and finding the source of life to bring you forth into a richer or mature condition" (Campbell, 1988, S. 152). Die Heldenreise, wie sie in der Folge beschrieben wird, umfasst die erforderliche Transformation eines Menschen zu einer erkenntnisreicheren[12] Version und ist symbolisch zu verstehen (Kolmar, 2021, S. 61). Die narrative Struktur einer inneren Heldenreise beginnt mit einer Berufung, einem Aufbruch: Der*die Held*in nimmt eine Reihe von Zeichen war, bis er*sie die Nachricht nicht unbeachtet lassen kann (Campbell, 1953, S. 56, 61). Dies kann zufällig wirken - grundsätzlich agiert ein Held aber aus der Position der Sicherheit (Kolmar, 2021, S. 216). Der Fokus liegt in dieser Phase auf dem Verlassen der bekannten Welt und dem Betreten einer unbekannten (vielleicht auch unbewussten) Welt. Diese unbekannte Welt wirkt sowohl anziehend als auch gefährlich (Campbell, 1953, S. 62, 83).

Bevor der*die Held*in die bekannte Welt verlässt, kann der Ruf auf taube Ohren (stossen) und die Antwort vorerst ausbleiben. Möglicherweise vergräbt sich der*die Heldi*n vorerst in Langeweile, Geschäftigkeit oder etwas anderes Kulturelles. Der*die Held*in erhält dann zum Teil mystische Hilfe von Mentor*innen. Verlässt der*die Held*in schliesslich die bekannte Welt, "kann dies einer Selbstvernichtung gleichkommen" (Campbell, 1953, S. 92-93).

Der*die Heldin verspürt durch die anfängliche Verunsicherung der eigenen Identität die psychologische Notwendigkeit zu klären 'was da ist'. Ein Narrativ des 'sich verstehen/sich verändern wollen' bringt den*die Held*in schliesslich auf die innere Heldenreise. In der unbekannten Welt, dem sogenannten Bauch des Walfisches, kämpft der*die Held*in gegen die eigenen inneren Dämonen/Drachen[13]. Hier wird dem*der Held*in erstmals die Grösse der Aufgabe bewusst. Darauf folgt eine Phase der Prüfungen und Beloh-

[8] Auf die Notwendigkeit sich selbst kennenzulernen wird im Kapitel der Heldenreise erneut eingegangen und wird dort im Detail geklärt.

[9] Ein anderes Beispiel stellt die Achtsamkeit dar: "Achtsamkeit ist eine buddhistische Bewusstseinstechnik der Selbstdistanzierung, die im weitesten Sinne Leiden vermindern soll. Im Management wird sie jedoch als Methode zur Produktivitäts- und Effizienzsteigerung bei hoher Arbeitsbelastung angewendet" (Girkinger, 2019, S. 47).

[10] Entsprechend bezeichnet Held*in in dieser Arbeit nicht notwendigerweise eine Person, die eine ausseralltägliche Leistung (oftmals im Krieg oder Kampf), eine Heldentat erbracht hat, wie dies beispielsweise im Duden als Bedeutung aufgeführt ist (Duden, 2023a). Zudem bezieht sich die Heldenreise nicht ausschliesslich auf das männliche Geschlecht. Das Heroische spielt an der Grenze zwischen Ordnung und Unordnung und umfasst sowohl Verteidigungen als auch Überschreitungen von Ordnungen (Kolmar, 2021, S. 62).

[11] Der Monomythos ist keine unumstrittene Theorie, die für diese Arbeit als Entwicklungsgeschichte verwendet wird. Für die Limitationen des Monomythos sei auf weiterführende Literatur verwiesen.

[12] Wobei sich Erkenntnis auf die Umwelt und die eigene Person/ das wahre Selbst bezieht.

[13] Beispielsweise der Verlust eines geliebten Menschen, der Umgang mit einer schwerwiegenden Krankheit sowie alle weiteren (Lebens-)Krisen (Herold, 2020). Ein alltäglicher Krisenmoment beginnt in dem "Moment, in dem man gewahr wird, dass man nicht anders kann, als Risiken einzugehen, um das Richtige zu tun" (Kolmar, 2021, S. 71).

nungen wie das Verspüren von Frieden, Klarheit, Neuheit, Glück, Grosszügigkeit, Einssein mit der Menschheit oder Teil von etwas (anderem) Grösserem zu sein (Campbell, 1953, S. 106).[14]

Nach anfänglichem Zögern in die bekannte Welt zurückzukehren, begibt sich der*die Held*in entweder aus innerem Antrieb, äusserer Rettung oder Zwang zurück in die Welt des Alltags.[15] Nach dem Übertritt in die bekannte Welt muss sich der*die Held*in mit der Integration des Erlebten (in der unbekannten Welt) in die bekannte Welt auseinandersetzen: Der* die Held*in vereint Alltagsleben mit dem neu gefundenen Wissen und damit die innere Welt mit den Erwartungen von Aussen. Durch das (Mit-)Teilen des Erlebten gewährt der*die Held*in den Menschen der bekannten Welt eine neue Freiheit (Herold, 2020). Abbildung 1 zeigt die visualisierte Heldenreise.

Die Transformität[16] der Heldenreise macht ihre Kommunikation so komplex. Sie kann nicht erzählt werden, sondern muss von jeder Person auf ihre eigene Art erlebt werden. Zudem soll hier auf zwei Paradoxa hingewiesen werden, die für den weiteren Verlauf der Arbeit von Bedeutung sind: Obwohl der*die Held*in auf der eigenen Reise die Grenzen zwischen bekannter und unbekannter Welt vorerst überwindet und schliesslich auflöst, kann der*die Held*in im grösseren Kontext sowohl Beschützer*in als auch Zerstörer*in einer grösseren Ordnung sein (Kolmar, 2021, S. 61)[17].

Kolmar argumentiert, dass Campbell's Selbsterkenntnis gewinnende Entwicklungsgeschichte eine "Konfrontation mit den eigenen Traumata darstellt" und "nur aus einer Position der prinzipiellen Sicherheit" (Kolmar, 2021, S. 216) funktionieren kann. Zentral für diese Sicherheit sind die spirituellen Helfer*innenfiguren[18]. Die Helden und Heldinnen konfrontieren ihre inneren Dämonen entsprechend nicht ohne Vorbereitung.

Heldentum ist in unserer Gesellschaft vielfach in die Medienwelt ausgelagert worden. Im Anhang A wird beschrieben, welche Transformationsarten Heldenforscher*innen in Filmen feststellen konnten und welche Funktionen Held*innen in unserer Gesellschaft zugeschrieben werden. Generell lässt sich festhalten, dass Held*innen die Gesellschaft in bedeutsame Weisen voranbringen. In anderen Worten erinnern uns Held*innen daran, dass das Leben selbst grösser ist als die Dimensionen, die wir annehmen mussten (Neiman, 2010). Im Folgenden wird auf diese für die Bachelorarbeit zentrale Funktion des Sinnempfindens eingegangen.

2.3.2. Heldenreise, Sinn und das Erhabene

Das Erfahren von Zugehörigkeit, Sinn sowie dem gelingenden Leben ist eng mit Erfahrungen des Erhabenen verbunden. Das Erhabene beschreibt die Grenzerfahrung zwischen der bekannten und der unbekannten Welt. Diese Grenzerfahrung ist eine wesentliche Komponente der Heldenreise. "Das Überschreiten der Ordnung, das Entdecken des Unbekannten in uns und in der Welt sind zentral für die Fülle des Lebens" (Kolmar, 2021, S. 217). Die Empfindung, an die Grenzen des eigenen Verständnisses zu stossen kann sich überwältigend und gleichwohl faszinierend anfühlen. Die erlebte Grenzerfahrung kann zudem tiefgreifende Veränderungen der Personen und ihrer Weltanschauungen nach sich tragen. In der Empirie wird als Sinnerfahrungen von transformativen Erfahrungen dieser Art gesprochen (Kolmar, 2021, S. 27). In der empirischen Forschung zu Sinnerfahrungen wird ein Schema sichtbar: "Es sind Erfahrungen, die nach Maslow etwas mit Selbsttranszendenz, mit Einssein und Zugehörigkeit in einem abstrakteren Sinne oder im Sinne der Gruppenzugehörigkeit zu tun haben" (Kolmar, 2021, S. 23).

Das Erhabene kann aus der Reise nach Innen erfahren werden, wenn über die bisherige Vorstellung des Selbst hinausgegangen wird und steht auch im Zentrum anderer Entwicklungsethiken (Kolmar, 2021, S. 81, 89–90). Hier sei erneut betont, dass das Erhabene von ambivalenter Natur ist. Einerseits kann das Erhabene eine Art von Furcht (Englisch 'terror') hervorrufen. Andererseits kann es eine tiefe Freude oder Ekstase auslösen (Englisch 'bliss'). Diese beiden Gefühle können abwechselnd oder zeitgleich auftreten und sind vielfach eng miteinander verbunden. In anderen Worten beschreibt das Erhabene "eine Grenzerfahrung, eine Situation existenzieller Unsicherheit, von der eine transformative Kraft ausgeht" (Kolmar, 2021, S. 81). Die Überforderungsmomente, die sich aus der Gefährdung der eigenen Existenz ergeben, können Ausgangspunkt für Verantwortungsübernahme und Moral sein (Kolmar, 2021, S. 90). In Hinsicht auf die Anpassungen im Handeln der Personen werden Ähnlichkeiten zur Tugendethik erkennbar. Transformative Erfahrungen sind Erfahrungen des Lernens, Hinterfragens und Entwickelns (Kolmar, 2021, S. 70).

Am Übergang zwischen Bekannt und Unbekannt wird "Freiheit und Zugehörigkeit spürbar". (Kolmar, 2021, S. 28). Das Erhabene birgt grosses Potential für positive Wachstums- und Befreiungserfahrungen. Die Heldenreise illustriert eine Möglichkeit, wie Transzendenz erreicht werden kann. "Sinn entsteht aus einem vertieften Verständnis des Menschsein,

[14] In den Sagen werden in dieser Phase oftmals auch Begegnungen mit Gottheiten oder anderen mythischen Wesen sowie sehr spezifische Inhalte wie Verführung der Frau und Versöhnung mit dem Vater beschrieben, worauf im Rahmen dieser Arbeit nicht weiter eingegangen werden kann.

[15] Oftmals ist in diesem Schritt in Mythen Magie involviert oder der*die Held*in profitiert von einem mitfühlenden Akt gegenüber einem vermeintlich 'niederen Wesen', den er oder sie auf dem Hinweg in die unbekannte Welt begangen (Campbell, 1953, S. 90 ff.).

[16] Transformative Erfahrungen beschreiben tiefgreifende Veränderungen, die oft schwer in Worte zu fassen sind. Dies einerseits, weil jede transformative Erfahrung individuell ist und andererseits die Sprache die Komplexität der Erfahrung kaum erfassen kann. Transformative Erfahrungen sind somit vergleichbar mit einem "Aufkleber auf einer Dose", der deren Inhalt beschreibt, obwohl der Inhalt nicht durch die Worte ausgedrückt werden kann (Kolmar, 2021, S. 22). Laut Aristoteles stellt der Zustand des höchsten menschlichen Wohlbefindens und Glücks, die sogenannte Eudaimonie, eine solche Erfahrung dar, die nicht in Worte gefasst werden kann und durch das Streben nach Tugend und der Entfaltung des eigenen Potentials erreicht wird. Auch in dieser Arbeit kann das Glücksempfinden eine transformative Erfahrung sein.

[17] Die Bedeutung dessen sollte spätestens im Kapitel zur äusseren Heldenreise geklärt werden.

[18] In den Mythen waren die Mentoren oft spirituelle Lehrer*innen. Heute leisten beispielsweise Therapeuten und Therapeutinnen die mentale und physische Vorbereitung (Kolmar, 2021, S. 216).

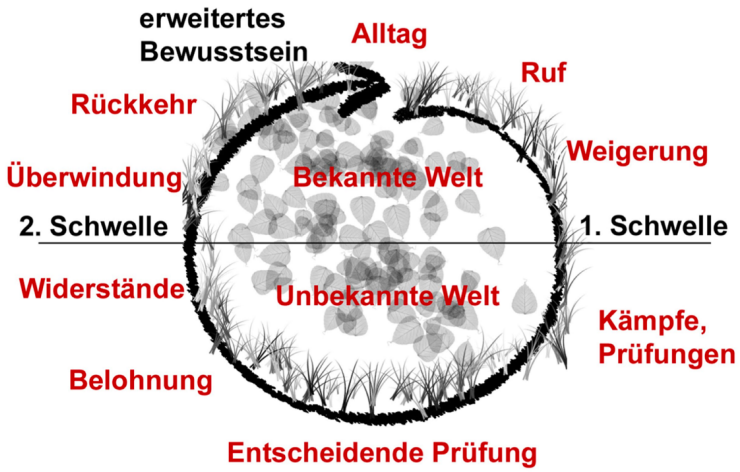


Figure 1: Held*innenreise nach Campbell (Kolmar, 2021, S. 123)

von dem die Heldenreise spricht" (Kolmar, 2021, S. 282) und folgt aus Haltungen, die man kultiviert. Abschliessend sei erneut darauf aufmerksam gemacht, dass Heldenreisen symbolisch zu sehen sind: Sie stellen Transformationen im Alltag dar, die es ermöglichen, eine Leere durch Sinnerfahrung zu überwinden. Im nächsten Kapitel wird beleuchtet, wie das Heldenhafte in der säkularen Kultur möglich oder eben problematisch ist.

2.3.3. Postheroische Gesellschaft, äussere Heldenreise und quasi Helden

Wissenschaftler*innen sprechen heutzutage von einer postheroischen Gesellschaft (Kolmar, 2021, S. 61). "Im Gegensatz zur heroischen Gesellschaft, die sich über die Erinnerung an ihre Gründerväter definiert, bezieht die postheroische Gesellschaft ihre Identität vorwiegend aus der Ablehnung von Heldentum und nationalem Opfermut" (Kolmar, 2020, S. 7). Heroische Bündnisse haben vielfach ein religiöses Zentrum, während postheroische Gesellschaften postheroische Aufträge fördern. "In heroischen Gesellschaften ist das Heldenbild eindeutig an eine moralische und metaphysische Ordnung zurückgebunden" (Kolmar, 2020, S. 12). Diese Ordnung wurde durch die europäische Aufklärung abgesetzt, was dem Heldenhaften den Sinn entzog. Eine heroische Orientierung ist in dieser Gesellschaftsform nicht unmöglich, aber problematisch, da Unordnung und Wirbel möglichst unterbunden werden sollten. Die postheroische Gesellschaft ist nicht von ausser- sondern von innerhalb des Menschen bedroht. Stabilität in der Ordnung wird durch die eigene Mässigung hergestellt und aufrechterhalten: Der*die moderne Held*in zeichnet sich durch seine*ihre Willens- und Triebskontrolle aus. Verzicht ist in der heutigen Gesellschaftsordnung heroisch geworden. Fantasien, Wünsche und Leidenschaften werden ungenügend - so zum Beispiel während eines Kinobesuchs, nicht aber im echten Leben - zugelassen. Das restliche Held*innenhafte ist in den kulturellen Konsum verdrängt. "Willenskontrolle als Triebverzicht ist Motor eines ruhelosen aber sterilen Daseins" (Kolmar, 2020, S. 25). Diese Rationalität richtet sich laut Adorno und Horkheimer gegen die Individuen selbst: Sie besiegen ihre eigene Natur, indem sie sie unterdrücken (Horkheimer & Adorno, 1987). "Es handelt sich um eine Zielerreichung ohne Ziel" (Kolmar, 2021, S. 63).[19]

An die Stelle des*r (prä-modernen) Held*in "trat die Idee des technokratischen Experten und des Staats als Versorgungsanstalt. Die Gesellschaft wird zu einer Maschine, die optimal gesteuert werden muss" (Kolmar, 2021, S. 62). In der postheroischen Gesellschaft ist möglichst viel geregelt, so dass sich das Individuum für korrektes Verhalten an den vorgefertigten Regelkatalog halten kann. Die Verantwortung verschiebt sich vom Individuum auf die System-Ebene[20]: Traditionelle Grundsätze und Tugenden werden durch Pragmatismus und Hedonismus ersetzt (Kolmar, 2021, S. 64-65). Die postheroische Persönlichkeit ist zudem im Gleichschritt mit der raschen sozialen Veränderung und somit ausgesprochen adaptiv. Zudem sind in der kapitalistischen Demokratie alle politisch gleich, was keinen Platz für individuelles Helden*innentum lässt. "Im postheroischen Gesellschaftsmodell sind Tausch, Konsum und

[19] Ein Beispiel stellt das Streben nach guten Noten in der Schule, mit dem Ziel an einer Universität studieren zu können, dar, welches schliesslich in das Streben nach guten Noten an der Universität, mit dem Ziel einen guten Beruf zu erhalten, übergeht. Im Beruf wird dann wiederum nach Leistung gestrebt, um eine Beförderung zu erhalten und so verschiebt sich das 'Ziel' weiter.

[20] Obschon beispielsweise politische, System-getreue Partizipationsmöglichkeiten bestehen. Oder in anderen Worten: "In der Praxis bedeutet Eigenverantwortung aber dann doch häufig nur, die staatlich gesetzten Regel einzuhalten, mehr nicht" (Kolmar, 2021).

Inszenierung die grundlegenden gesellschaftlichen und individuellen Wahrnehmungskategorien" (Kolmar, 2021, S. 64). Das Individuum wird mit dem Glücksversprechen sowie der Leitidee "sei du selbst" und "mach etwas aus deinem Leben" zurückgelassen. Schlussendlich bezieht das Individuum die kapitalistische Tausch- und Konsum-Logik auf sich selbst und versteht sich als Kapital[21].

In der säkularen Welt haben sich äussere kapitalistische und anderweitig säkulare Heldenreise herausgebildet. Diese machen den Anschein dieses Transformations- und Lernprozesses, sind es aber nicht. Äussere Heldenreisen folgen der gleichen Struktur wie innere Heldenreisen, haben etwas Transformatives, lassen jedoch keine Ordnungsüberschreitung zu. Im demokratischen System ist der*die Held*in "einer Vernunft unterworfen, die ihm*ihr ständig Regeln auferlegt" (Kolmar, 2021, S. 69). Der*die Held*in muss in einem komplexen System zwischen Ordnungsüberquerung und Ordnungsstabilisierung navigieren. In dieser Bachelorarbeit wird davon ausgegangen, dass Selbstoptimierung ebenfalls eine äussere Heldenreise darstellen kann.

Grundsätzlich lässt sich fast jede individuelle Geschichte als Heldenreise darstellen. Folgende Beispiele illustrieren äussere Heldenreisen, die von sogenannten Quasi Helden gegangen werden. Ein*e Aktivist*in, die sich auf alle möglichen Weisen, vom auf-die-Strasse-kleben bis zum Gewalteinsatz, gegen etwas auflehnt, befindet sich auf einer äusseren Heldenreise. "Der Akt des sich Auflehnens ist an sich noch nicht heroisch, er ersetzt die Frage nach dem Sinn nicht und wird zur blossen Attitüde. "Im Gegenteil kann man argumentieren, dass ein sinnloses Aufbegehren gegen 'das Sterben des Lichts' ein Mangel an Mut und Einsicht ist, das Unabweisbare anzuerkennen" (Kolmar, 2021, S. 68). Auch die Rekrutierung von Kämpfer*innen für scheinbar heilige Kriege folgt der Struktur einer äusseren Heldenreise. Der IS beispielsweise verspricht ein Abenteuer in einer unbekannten Welt und einen Kampf für eine grössere Sache[22].

Der amerikanische Traum des sich vom-Tellerwäscherzum-Millionär-Hocharbeitens stellt die klassische säkulare Heldenreise dar. Heutzutage haben aufgrund der geringen sozialen Mobilität nur wenige die Möglichkeit, die kapitalistische Heldenreise in dieser extremen Form anzutreten. Eine abgeschwächte Form stellt schon das Streben nach guten Noten fürs Studium, gefolgt vom Streben nach guten Noten im Studium für einen guten Job, wiederum gefolgt vom immerwährenden Streben nach der Beförderung im Job, dar. Bei der Betäubung durch Beschäftigung handelt es sich um wiederum um die vorgängig beschriebene "Zielerreichung ohne Ziel" (Kolmar, 2020, S. 26).

Laut Hirschman (1977) befürworteten Theoretiker anfänglich kapitalistische Heldenreisen. "Sie sahen im Kapitalismus eine Möglichkeit, die 'dunkleren' Leidenschaften des Menschen nach Macht, Status und Gewalt in einer neuen symbolischen Ordnung so zu bündeln, dass ihnen nicht nur das direkt destruktive Potential genommen wird, sondern dass diese Leidenschaften vielmehr so kanalisiert werden, dass sie etwas gesellschaftlich Nützliches erreichen" (Hirschman, 1977, S. 134).

Die Auffassung der Heldenreise als Selbstbefreiung, während die Ordnung nicht verlassen werden kann, verdeutlicht ihre Ambivalenz. Wie im Kapitel zum säkularen Zeitalter beschrieben, verschwinden laut Taylor die menschlichen Bedürfnisse mit ihrer Unterdrückung nicht. Der*die Held*in der äusseren Heldenreise wird innerhalb der Ordnung nie in einem befriedigenden Masse über sich hinausgehen, tiefe Verbundenheit mit anderen Menschen, Lebewesen oder dem Universum verspüren und so um die unbeschreibbare Mitte kreisen. Entsprechend kann heute nicht mehr von einer produktiver Zügelung dunkler Leidenschaften gesprochen werden. Es besteht die Gefahr, dass die äussere Heldenreise als Weg zum Glück gesehen wird, dies aber, im Gegensatz zur inneren Heldenreise, nicht herbeiführen kann. Im nachfolgenden Kapitel wird unter anderem auf diese Problematik in einer Zwischenbilanz von Heldenreisen und Selbstoptimierung eingegangen.

2.4. Zwischenbilanz Heldenreisen und Selbstoptimierung

Sowohl Heldenreisen als auch Selbstoptimierungsnarrative gehen von einem defizitären Individuum aus. Beide Phänomene beschreiben eine Transformation des Individuums. Während der Veränderung ist deren Ausgang unklar – es besteht keine Erfolgsgarantie und möglicherweise Widerstand. Personen, die sich selbstoptimieren oder sich auf einer Heldenreise befinden, können ein Bewusstsein für ihre Transformation haben.

Die innere Heldenreise geht vielfach mit einer starken Verbundenheit mit der Gesellschaft einher. Das soziale Umfeld spielt für die Transformation des*der Helden*in eine zentrale Rolle: Helden und Heldinnen sind von zahlreichen Freunden, Geliebten und insbesondere Mentoren und Mentorinnen umgeben. Diese unterstützen den*die Heldin durch ihre tatsächliche, imaginäre oder implizite Anwesenheit auf der Reise (Campbell, 1949). Selbstoptimierungspraktiken können auch als Singularisierung[23] verstanden werden.

Während bei der Selbstoptimierung die eigenen Grenzen möglicherweise erweitert werden, findet bei der inneren Heldenreise eine echte Grenzüberwindung statt. Die Selbstoptimierung und andere äussere Heldenreisen finden in der bestehenden epistemischen Ordnung[24] statt: Die Individuen halten mit der kapitalistischen Wachstumsdynamik Schritt

[21] Es sei erneut auf das Kapitel zur Selbstinvestition im Anhang A verwiesen.

[22] Auf Gruppendynamiken der teils radikalen Gruppierungen kann im Rahmen dieser Arbeit nicht eingegangen werden, obwohl gewisse Elemente scheinbar heroisch sind. Es sei unteranderem auf (Münkler, 2011, 2015) verwiesen.

[23] Reckwitz (2017) bezeichnet den sozialen Prozess, durch den individuelle Lebensgestaltungen und Identitäten zunehmend von traditionellen gemeinschaftlichen Normen entkoppelt und diversifiziert werden, als Singularisierung.

[24] Mit epistemischer Ordnung sind hier die kapitalistischen kulturell geprägten Narrative von Mangel und Konsum gemeint, anhand deren wie vorab beschrieben dunkle Leidenschaften des Menschen gezähmt werden sollten.

und versuchen durchs schneller, besser, schöner werden die innere Leere zu füllen (Lacan, 1977, 1992). Das Inkorporieren dieser Logik stellt Teil des Problems und nicht Teil der Lösung dar – dem eigentlichen Problem wird aus dem Weg gegangen. Denn die eigenen Monster werden nicht konfrontiert, sondern es werden neue geschaffen: ein erlebter Mangel an Zuneigung kann beispielsweise mit einer Schönheitsoperation zu heilen versucht werden. Dabei wird der Mangel nicht direkt adressiert und das defizitäre Selbstbild womöglich verstärkt. Das Scheitern des Vorhabens ist vorprogrammiert und was sich wie eine Selbstbefreiung anfühlt, ist letztlich keine. Wahrer Mut bestände darin, die Welt so zu akzeptieren, wie sie wirklich ist. Die unbekannte Welt kann darin bestehen, von dominanten kulturellen Praktiken abzuweichen.

Selbstoptimierung kann, aufgrund der beschriebenen Unterschiede, nicht an die Tiefe der Selbst- und Menschenkenntnis herankommen, die durch eine innere Heldenreise erreicht werden kann und in direkter Verbindung zu Glück steht (Kolmar, 2021, S. 73, 75). Fehlt den Individuen das Verständnis bezüglich dieser Unmöglichkeit, können psychische Krisen und Krankheiten wie Burnout oder Depressionen ausgelöst werden.

Entsprechend ist es in dieser Arbeit von zentralem Interesse herauszufinden, welches Verständnis die Befragten von ihren persönlichen Entwicklungsreisen haben und was sie sich davon erhoffen. Ein Modell aus der positiven Psychologie, welches langfristiges Wohlbefinden erfasst, wird in dieser Arbeit als Annäherung an ein gelingendes Leben verwendet und im nächsten Kapitel beschrieben.

2.5. Das gelingende Leben anhand des PERMA-Modells

Die Positive Psychologie ist eine Strömung der Psychologie, die insbesondere durch die Psychologen Martin Seligman und Csikszentmihalyi Ende des 20. Jahrhunderts berühmt geworden ist. Im Gegensatz zur defizitorientierten traditionellen Psychologie, untersucht die Positive Psychologie, was Menschen zum Aufblühen bringt. Bei der Positiven Psychologie handelt es sich um eine tugendethisch- und Entwicklungs-orientierte Theorie des gelingenden Lebens: Im Fokus steht die Erarbeitung positiver Gewohnheiten, wozu ein hohes Mass an Selbstkenntnis notwendig ist (Kolmar, 2021, S. 58).

Als Teil der empirisch orientierten Forschung zu menschlichem Glück, ausgedrückt als persönliche Lebenszufriedenheit, entwickelte Seligman das sogenannte PERMA-Modell. Fünf operationalisierte und messbare Bereiche konstruieren entsprechend Seligman das menschliche Wohlbefinden. PERMA ist ein Akronym für die Begriffe Positive Emotionen (für Positive Emotions), Engagement, Beziehungen (für Relationships), Sinn und Zweck (für Meaning and Purpose) und Errungenschaften (für Accomplishments), die gemäss Seligman, richtig kombiniert ein erfülltes Leben ermöglichen. "Mit der an 'permanent' angelehnten Benennung PERMA spielt Seligman auf den Aspekt der Nachhaltigkeit des so erreichten Wohlbefindens an" (Blickhan, 2017, S. 42). Basierend auf dem PERMA-Modell wurde von Butler und Kern der PERMA-Profiler[25] entwickelt. Nachfolgend werden die einzelnen PERMA-Kategorien beschrieben.

Zu den Positiven Gefühlen gehören positiv wahrgenommene Sinneserfahrungen wie Freude, Spass, Befriedigung oder Faszination. Genusserfahrungen können nach der 'Broaden and Built Theory' Ausgangspunkt für Persönlichkeitsentwicklungen sein (Fredrickson, 2001). Im Gegensatz zu Buddhistischen oder stoischen Ansätzen, die das Glück in möglichst konsequentem Ablösen von äusseren Anhaftungen sehen, wird hedonischen Ansätze in der Positiven Psychologie ein Beitrag zum persönlichen Glück zugeschrieben (Haidt, 2006, S. 90). Mit einem Blick auf die anderen PERMA-Bereiche wird schnell klar, dass ein Leben in Saus und Braus, ohne Zugehörigkeitsgefühl und das Erfahren von Sinn, nicht als erfüllend erlebt wird.

Engagement bezieht sich in der Positiven Psychologie stark auf sogenannte 'Flow' Momente, in welchen eine Person so vertieft in eine herausfordernde aber mit den eigenen Kompetenzen bewältigbare Aufgabe ist, dass sich die Person in einen Zustand der Selbstvergessenheit befindet. Csikszentmihalyi fand heraus, dass eine Mehrheit Momente des 'Flows' über Genussmomente wie Schokolade essen oder Sex bevorzugt (Haidt, 2006, S. 95) und dass diese 'Flow' Momente am häufigsten bei der Arbeit und der freien Zeit erlebt werden (Csikszentmihalhi, 2020).

Positive Beziehungserfahrungen sind essenziell fürs Wohlbefinden. Zahlreiche Studien zeigen, dass ein unterstützendes soziales Netzwerk beispielsweise mit höherer Gesundheit, weniger Depression und geringerer Sterbewahrscheinlichkeit korreliert (S. E. Taylor, 2011).

Einen Sinn und Zweck im Leben zu haben, wird oft als das Gefühl, Teil eines Grösseren zu sein, beschrieben. Einen Sinn in diversen Tätigkeiten im Leben zu verspüren oder sich von seinem Umfeld geschätzt fühlen, geben ebenso Sinn und Zweck (Steger, 2013). Sinn vermittelt das Gefühl, dass das eigene Leben bedeutend ist. Studien zeigen, dass dieses Gefühl mit höherer Lebenszufriedenheit und besserer Gesundheit einhergeht (Boyle et al., 2009; Ryff et al., 2004; Steger, 2013).

Errungenschaften werden subjektiv unterschiedlich wahrgenommen. Es lässt sich festhalten, dass ein Empfinden, auf Ziele hinzuarbeiten und sie zu erreichen, zu Errungenschaften gehört. Zudem sind gewisse Kompetenzen und das effektive Erreichen von Zielen und Aufgaben charakteristisch (Seligman, 2018).

Alle Bereiche bauen auf den individuellen Stärken der Personen auf. Die eigenen Stärken zu kennen und diese einsetzen zu können trägt durch die positive Beeinflussung aller

[25] Der Profiler wurde an mehreren 1'000 Personen getestet und besteht aus insgesamt 15 Aussagen, die alle Bereiche des PERMA-Modells mit je drei Aussagen erfassen. Die Befragten evaluieren auf einer Skala von 0 bis 10, inwiefern die Aussage auf sie zutrifft. Die Statements enthalten ausschliesslich positiv konnotierte Wörter (Butler & Kern, 2016, S. 9). Laut den Erfindern hat der Fragebogen eine beschreibende Natur. Es existiert bislang kein Idealprofil. Allerdings können mit den entsprechenden Strategien auffällig tiefe Werte in einem der PERMA-Bereiche adressiert und verbessert werden. Auf den PERMA-Profiler und dessen Verwendung wird erneut im Detail im Methodikteil dieser Arbeit eingegangen.

PERMA-Bereiche positiv zum individuellen Wohlbefinden bei (Peterson & Seligman, 2012). Anderweitig ebnen ausgewogene Handlungen in sowie Widerspruchslosigkeit zwischen den Bereichen laut Seligman den Weg zum erfüllten Leben.

Während beispielsweise mittels Konsum in der modernen Gesellschaft nach Glück gestrebt wird, wird das Empfinden von Sinn in dieser problematischer. Aus der Theorie zu den Heldenreisen geht hervor, dass die Transformation innerer Heldenreisen dem gelingenden Leben zuträglich ist. Quasi Heldenreisen wie die Selbstoptimierung eine darstellt, können scheinbar zu Glück führen. In dieser explorativen Arbeit soll herausgefunden werden, welche Selbstoptimierungspraktiken und -narrative für Studierende der Wirtschaftswissenschaften existieren, und ob diese in Konkurrenz mit Heldenreisen stehen. Die Ergebnisse sollen im Einbezug der vorherrschenden Vorstellungen eines gelingenden Lebens sowie dem Gesellschaftsbild betrachtet werden. Die Vorgehensweise dazu ist im nächsten Kapitel erläutert.

3. Methoden

3.1. Auswahl

3.1.1. Stichprobe

Die Forschungsfrage dieser Bachelorarbeit ist auf Studierende der Wirtschaftswissenschaften beschränkt. Dies einerseits, weil die 'Verwirtschaftlichung' des Sozialen, wie sie von Kritiker*innen beschrieben wird, besonders bei Studierenden dieser Fachrichtung erwartet werden kann und andererseits, weil die Interviewerin selbst an der Wirtschaftsuniversität St. Gallen studiert und ein entsprechendes Netzwerk an der Universität vorweisen kann. Mögliche Studiengänge der Befragten sind alle Bachelor- und Masterstudiengänge der Universität St. Gallen[26] sowie Wirtschaft im Neben- oder Hauptfach an anderen Schweizer Universitäten. Die Befragten sollten zudem zwischen 18 und 32 Jahren alt sein. Dies in der Annahme, dass die Lebensrealitäten von Bachelor oder Masterstudierenden in diesem Alter nicht massgeblich voneinander abweichen, was andernfalls zu unerklärbaren Unterschieden in den Beantwortungen hätte führen können[27]. Eine weitere Voraussetzung war, dass die Befragten Deutsch[28] sprechen, da die Befragung auf Deutsch durchgeführt wurde.

3.1.2. Kombination qualitativer und quantitativer Methodik

Um mögliche Zusammenhänge zwischen den Phänomenen Selbstoptimierung, Heldenreisen sowie Vorstellungen eines gelingenden Lebens und der Rolle von Sinn darin zu erkunden, wurde aufgrund der Intimität der Themen ein persönliches Gespräch als geeignet beurteilt. Da das subjektive Denken, Handeln, Fühlen und die Vorstellungen der einzelnen Personen im Fokus stehen, wurden individuelle Interviews als erste Datenerhebungsform auserkoren[29]. Bei der Auswahl der Interviewform war entscheidend, dass die Befragten ein Format haben, in welchem sie persönliche Inhalte erzählen können und die Befragerin die Möglichkeit hat bei Unklarheiten nachzuhaken.

Die Wahl fiel entsprechend auf ein semistrukturiertes[30] Interview mit Elementen eines narrativen Interviews. Narrative Interviews werden verwendet, um erlebte Erfahrungen zu erfragen (Aghamanoukjan et al., 2009, S. 421). Entsprechend bestehen Ähnlichkeiten zu persönlichen Praktiken und Rechtfertigungsnarrativen. Je nach Strukturierungsgrad (voll- oder teilstrukturiert), gibt es mehr oder weniger geschlossene Fragen mit vorgegebenen Antwortmöglichkeiten, die in einer definierten Abfolge gestellt werden (Diekmann, 2007, S. 374). Der Ausgestaltung des Interviews ist ein separates Kapitel gewidmet.

Basierend auf der Vermutung, dass die Interviewten in den Interviews nicht die ganze Breite der explorativen Fragen beantworten würden, stellt eine Methoden-Triangulation ein für diese Arbeit angemessenes Datenerhebungsvorgehen dar. Durch die Erstellung eines Fragebogens mit vorstandardisierten und sowie selbstreflexiven Fragen, den die Interviewten im Anschluss an das Interview ausfüllen sollten, kann die ganze Breite der Fragen sowie deren Wirkung erfragt werden. Der Fragebogen komplementiert die Interviews[31]. Laut Kelle und Erzberger werden in der Forschungspraxis zur gegenseitigen Ergänzung vermehrt quantitative mit qualitativen Methoden zu einem Forschungsdesign kombiniert (Flick & Kardorff, 2006, S. 300-303)[32]. Mit anderen Worten wurde in dieser Arbeit induktiv und deduktiv gearbeitet, um dasselbe soziale Phänomen, Selbstoptimierungsnarrative, zu erfassen: Bereits vorhandene Theorie wurde bei der Erstellung und Kodierung der Interviews sowie der Erstellung des Fragebogens einbezogen. Gleichzeitig bestand eine Offenheit, bei der Auswertung neue Kodes hinzuzufügen und neue Thesen aus der

[26] Beispielsweise Bachelor in Volkswirtschafts- oder Betriebswirtschaftslehre, Bachelor oder Master in Internationalen Beziehungen, Wirtschaftspsychologie oder Master in 'Business Innovation'.

[27] Ein Leben veränderndes Beispiel stellt die Geburt eines Kindes oder ähnliches dar. In dieser Arbeit wurde keine spezifische Generation in den Fokus genommen.

[28] Muttersprache oder mind. C1-Niveau.

[29] Diekmann (2007, S. 186 ff.) beschreibt, dass nicht jede Methode für jedes Forschungsziel geeignet ist. Insbesondere für explorative Studien ergeben sich präzise Fragen vielfach erst im Verlauf der Untersuchungen. Die Kombination aus einem Interview, in welchem persönliche Wahrnehmungen erfasst werden können sowie einem Fragebogen, welcher ein gewisses Mass an Standardisierung in die Forschung bringt, wurde für das Mass an Explorativität dieser Studie als angemessenes Forschungsdesign empfunden.

[30] Synonym für semistrukturiert wird in der Literatur auch von halbstandardisiert und halbstrukturiert gesprochen. Es würde den Rahmen dieser Bachelorarbeit sprengen, auf andere Interviewformen einzugehen. Als weiterführende Literatur seien (Diekmann, 2007) sowie (Flick & Kardorff, 2006) empfohlen.

[31] Die Daten der Interviews und des Fragebogens sind gleichwertig. Die Daten des Fragebogens können standardisierter ausgewertet werden, während die Daten aus den Interviews tiefer gehen und die Daten des Fragebogens kontextualisieren können.

[32] Zudem kann durch die Methodenkombination einer Scheinsicherheit der Daten entgegengewirkt werden. Interviews und Fragebogen dienen sich gegenseitig als Interpretationshilfen.

empirischen Forschung aufzustellen. Bevor die ineinandergreifende Auswertung von Interview und Fragebogen illustriert wird, wird in den nächsten Kapiteln die Erstellung, das Testen und die Durchführung der Interviews und dem Fragebogen beschrieben.

3.1.3. Interview

Inspiriert von den Erzählmöglichkeiten des narrativen Interviews, welches vielfach mit einer oder wenigen Fragen grosse Teile der Lebensgeschichten der Befragten erforscht, wurde ein Interview mit mehreren thematisch abgegrenzten Erzählblöcken entworfen. Die Befragten sollten so die Möglichkeit haben, themenspezifisch von ihren Wahrnehmungen, Vorstellungen und Lebenspraktiken zu erzählen. Die meisten Fragen wurden so breit gestellt, dass die Befragten die Chance haben sollten, ihre persönlichen Geschichten zu erzählen und eigenen Schwerpunkte zu setzen. Insbesondere weil bei heiklen und unangenehmen Fragen sowie bei Fragen, bei welchen starke soziale Erwartungen mitschwingen, mit Verzerrungen in Interviews zu rechnen ist (Diekmann, 2007, S. 441-443) wurde auf eine neutrale Interviewführung geachtet[33]: Dabei wird versucht, "der befragten Person das Gefühl zu vermitteln, dass ihre Antwortreaktionen weder in der Interviewsituation noch später in irgendeiner Weise sanktioniert werden. Die meist angekündigte Vereinbarung über die Vertraulichkeit der Angaben und eine neutrale Gesprächsführung sollen diesen Eindruck verstärken" (Diekmann, 2007, S. 439). Entsprechend wurde eine Vertraulichkeitserklärung in die Intervieweinführung eingebaut. Zudem wurden als Gegenmassnahme wie von Diekmann (2007, S. 449) empfohlen alle Fragen möglichst neutral formuliert und auf wertbesetzte Begriffe verzichtet. Im 'Briefing' wurde zudem erklärt, um was es im Interview geht und wie es abläuft (Flick & Kardorff, 2006, S. 367). Die finalen Interviewfragen können im Anhang B eingesehen werden.

Im ersten Themenblock wurden die Befragten zur Wahrnehmung der und dem eigenen Wohlbefinden in der Schweizer Gesellschaft gefragt. Ziel war es herauszufinden, ob die Befragten Entwicklungsmöglichkeiten und Sinn in der Gesellschaft verspüren. Darüber hinaus wurde untersucht, ob die Befragten ein postheroisches oder heroisches Gesellschaftsbild zeichnen. Die Überleitung zu den Selbstoptimierungsnarrativen wurde mit einer Frage zu den gesellschaftlichen Erwartungen gestartet. Danach wurden die individuellen Haltungen zur Selbstoptimierung erfragt. Es wurde angestrebt herauszufinden, welche Leistungserwartungen wahrgenommen, sich selbst auferlegt oder internalisiert werden.

Die Anschlussfrage sollte den Befragten den Anstoss geben, von ihren Selbstoptimierungspraktiken zu erzählen. Um die Rechtfertigungsnarrative für die Selbstoptimierung der Studierenden zu ermitteln, wurden Fragen zur Motivation gestellt. Im Interview wird sodann ein Themenwechsel angekündet. Zwei Fragen sollen die Befragten anregen, von ihren Vorstellungen eines gelingenden Lebens zu sprechen. Um mögliche Zusammenhänge oder Kompetitionsverhältnisse der Bereiche Selbstoptimierung und gelingendem Leben analysieren zu können, wurde die Idee des Arbeitens für ein gelingendes Leben in zwei Fragen erfragt. Die Studierenden sollten konkret nach persönlichen Zusammenhängen gefragt werden, damit in sich überschneidende Antworten nicht zu viel hineininterpretiert wird, sondern in Erfahrung gebracht werden kann, welche Verbindungen die Studierenden selbst sehen. Das Interview wird mit einer Ergänzungswunschfrage abgeschlossen.

Die finale Formulierung sowie die Reihenfolge der Frage wurden in einem Interviewleitfaden festgehalten. "Ein Leitfaden dient unter anderem dem Zweck, dass im Interview möglichst alle relevanten Aspekte und Themen angesprochen werden und damit eine gewisse Vergleichbarkeit der Antwortreaktionen verschiedener Befragter ermöglicht wird" (Diekmann, 2007, S. 537). Die Struktur des Interviews liess der Interviewerin die Möglichkeit, eine Frage in anderen Worten zu wiederholen oder auf andere Weise nachzufragen, sofern die Befragten die Frage unvollständig beantworteten oder falsch verstehen würden. Zudem wurden im Interviewprotokoll[34] Verhaltensregeln für den Ablauf des Interviews festgelegt, um ein möglichst positives Interviewklima zu garantieren.[35]

Die Interviewfragen wurden so konstruiert, dass durch die spontane Beantwortung die intuitiven Selbstoptimierungsverständnisse und -praktiken der Studierenden erfasst werden. Im Interview wurden die Studierenden absichtlich in keine Richtung gelenkt, während im Fragebogen die Antworten basierend auf vorhandener Theorie vorstrukturiert wurden.

Für das Interviewdesign wurde zudem recherchiert, welchen Effekt der Ort der Durchführung beziehungsweise der Kommunikationskanal auf die Ergebnisse hat. Laut Diekmann (2007, S. 501 f.) sind Antworten aus telefonisch durchgeführten Interviews und von Angesicht zu Angesicht durchgeführte Interviews vergleichbar. Aufgrund des anzunehmend geringen Qualitätsunterschieds sowie Praktikabilitäts- und Effizienzgründen[36] wurde entschieden, das Interview nicht in Person, sondern virtuell durchzuführen.

Bei den Interviews sowie beim Fragebogen, dem das nächste Kapitel gewidmet ist, wurde entsprechend Diekmann's Empfehlungen (2007, S. 479 ff.) bei der Frageformulierung darauf geachtet, kurze, verständliche und hinreichend präzise Formulierungen zu machen. Für die Interviews sowie für den Fragebogen wurde einfaches Hoch-

[33] Auf die neutrale Interviewführung wird im Kapitel zur Interviewdurchführung erneut eingegangen.

[34] Mit folgendem Link können das Interviewprotokoll sowie der finale Interviewleitfaden eingesehen werden.

[35] Worauf im Kapitel zur Interviewdurchführung erneut eingegangen wird.

[36] Zum einen fielen die Interviewtermine zur Hälfte in die Semesterferien, während deren die Studierenden nicht an ihren Universitäten anzutreffen gewesen wären. Zum anderen befand sich die Interviewerin nicht mehr im universitären Alltag. Letztlich variierte die Aufenthaltsorte der Studierenden zu stark, um effizient Interviews vor Ort zu koordinieren.

deutsch verwendet[37] und keine doppelten Verneinungen eingebaut, damit die Befragten nicht unnötig lange ins Nachdenken oder Stocken geraten. Zudem wurden keine mehrdimensionalen Fragen gestellt.

3.1.4. Fragebogen

Für die Konstruktion des Fragebogens wurden diverse bestehende Fragebogen und Theorien hinzugezogen: Die Fragen zum Gesellschaftsbild wurden massgeblich vom Fragebogen der 'Moral Foundations Theory' inspiriert, mit welcher der Psychologe Jonathan Haidth und seine Kollegen*innen basierend auf sechs definierten moralischen Grundlagen menschlichen Verhaltens zuverlässig politische Ideologien ausmachen können.[38]

Anstelle der moralischen Grundlagen wurden drei Achsen für das Gesellschaftsbild festgelegt: Sinn, Entwicklungsmöglichkeiten sowie Leistungserwartung. Wie bei der 'Moral Foundation Theory' wurden mehrere Aussagen pro Achse entworfen, denen die Befragten im Fragebogen dieser Bachelorarbeit auf einer siebenstufigen Likert-Skala[39] zustimmen können. Die Sinn-Achse sollte Auskunft darüber geben, wie säkular die Gesellschaft wahrgenommen wird. Der Entwicklungs-Achse wurden Heldenreisen ähnliche Aussagen zugeschrieben. Die Entwicklungsachse sollte messen, inwiefern die Befragten Möglichkeiten haben, sich selbst kennenzulernen und sich auf eine Helden*innenreise zu begeben. Die Leistungsachse dient speziell zur Erfassung des Verhältnisses zwischen Sinn- und Entwicklungsmöglichkeiten auf der einen und Leistungserwarten auf der anderen Seite. Die Formulierungen der Aussagen entspringen der gelesenen Theorie. Für die Sinn-Achse wurde für die Formulierung der Aussagen zudem der 'Meaning in life' Fragebogen[40] hinzugezogen. Alle Aussagen im Zusammenhang mit den Achsen sind angelehnt an die genannten Fragebögen und entsprechend in der 'Ich-Form' verfasst. Um eine Personenunabhängige Aussage zum gesellschaftlichen Glück, der Gesundheit und der Säkularität zu erhalten, wurden drei Kontrollfragen hinzugefügt, die neutral formuliert sind.

Zur Messung der Zustimmung wurden im Fragebogen durchgängig 7-stufige Likert-Skalen verwendet, denn "empirisch sprechen die Ergebnisse eher für ungerade Skalen" (Franzen, 2019, S. 670). Laut Moors (2008) konnten zwischen 5-stufigen und 6-stufigen Antwortmöglichkeiten keine signifikanten Unterscheide festgestellt werden. Obwohl das Risiko besteht, dass die Befragten vermehrt zur Mitte tendieren, wenn eine solche in den Antwortkategorien angeboten wird, wurde eine neutrale Antwortmöglichkeit angeboten, damit sich die Studierenden nicht in eine Richtung entscheiden müssen, obwohl sie einem Thema neutral gegenüberstehen. Weil zusätzlich bei Fragen, die sich nicht auf die eigene Person beziehen, eine 'ich weiss nicht' Kategorie am rechten Rand hinzugefügt wurde und den Studierenden bei fast allen Fragen die Möglichkeit gegeben wurde sich zu enthalten, sollte dieses Risiko minimiert werden[41].

Für die Erfassung der Selbstoptimierungspraktiken wurden zwei in der Theorie präsente Bereiche ausgewählt, das Aussehen sowie die Gesundheit, die analysiert werden sollen. Da es sich bei den Befragten um Studierende handelt, wurde zudem 'geistige Leistungsfähigkeit' als Bereich hinzugefügt. Die Bereiche 'soziale Beziehungen' und 'Glück' wurden hinzugefügt um zu prüfen, inwiefern die Studierenden diese Bereiche als aktives Arbeiten verstehen, und damit wichtige Hinweise für die Beantwortung der Forschungsfrage zu erhalten. Pro Bereich wurden in einem Matrixformat sechs bis neun Selbstoptimierungstätigkeiten aufgeführt und die Studierenden wurden gebeten, für jede Tätigkeit anzugeben, wie intensiv sie diese ausführen. Im Anschluss an jeden Tätigkeitsblock pro Bereich folgte ein Block, der die Gründe der Arbeit an sich selbst abfragte. Die Gründe stammen zum Grossteil aus der Literatur, dienten zur Erforschung der Rechtfertigungsnarrative und blieben über alle Bereiche konstant, während die Tätigkeitsbeispiele auf die Bereiche angepasst wurden. Sowohl bei den Tätigkeiten als auch bei den Gründen hatten die Studierenden die Möglichkeit, unter der Antwortmöglichkeit 'Andere' persönliche Beispiele hinzuzufügen.

Die Antwortmöglichkeiten zur Frage nach einem gelingenden Leben wurden vom PERMA-Profiler inspiriert. Dieser misst mit je drei Aussagen zu jeder PERMA-Kategorie menschliches Wohlbefinden (Immer, 2017). Die Antwortmöglichkeiten des PERMA-Profilers wurden im Rahmen dieser Bachelorarbeit als Vorstellungen eines gelingenden Lebens umformuliert. Entsprechend wurde der Wortlaut der Fragen zum Teil übernommen und auf ein gelingendes Leben, statt das Hier und Jetzt bezogen. Da nicht alle Fragen auf ein gelingendes Leben übersetzt werden konnten, wurden lediglich zwei Antwortmöglichkeiten pro PERMA-Kategorie gewählt sowie zwei Antwortmöglichkeiten zum Heldenreisen-ähnlichen 'sich selbst kennenlernen und weiterentwickeln' hinzugefügt.

Nach den Pretests wurde der Fragebogen mit zwei qualitativen, selbstreflexiven Fragen[42] ergänzt. Entsprechend

[37] Beim Fragebogen bietet sich Hochdeutsch als verwendete Schriftsprache im deutschsprachigen Raum an. Trotz möglichen Einbussen in der Lockerheit und Familiarität der Befragten wurde entschieden, die Interviews ebenfalls auf Hochdeutsch durchzuführen. Dies insbesondere zu Konsistenz- und Effizienzzwecken in der Auswertung.

[38] Der Moral Foundations Fragebogen, dessen Übersetzungen und Auswertungsvorlagen können auf der Moral Foundations Webseite eingesehen werden ("Questionnaires | Moral Foundations Theory", n.d.).

[39] Es gab je drei Zustimmungs- und Ablehnungsantwortmöglichkeit, eine neutrale Mitte sowie die Möglichkeit sich mit 'Ich weiss nicht' ganz zu enthalten.

[40] Es wurde die deutsche Übersetzung des 'Meaning in Life' Fragebogen aus der Arbeit von Neumüller (2013, S. 85) als Inspiration verwendet. Der' Meaning in Life' Fragebogen ist Messinstrument zum wahrgenommenen Sinn im Leben.

[41] Für die Auswertung wurde der neutralen Haltung in der Mitte der numerische Wert in der Mitte der 7-stufigen Likert Skala zugewiesen, während Enthaltungen oder die 'Ich weiss nicht' Antwort 0 Punkte erhielten, sprich aus der Auswertung herausfielen.

[42] Konkret, wie die Befragten die Beantwortung der Fragen zu den Selbstoptimierungstätigkeiten sowie dem gelingenden Leben erlebten. Für Details zum Pretest sei auf das Kapitel 3.2 verwiesen.

des zu haltenden Spannungsbogens wurden in den Interviews und im Fragebogen die wichtigsten Fragen im zweiten Drittel gestellt. Laut Diekmann (2007, S. 484-485) sollten sozialstatistische Fragen am Ende der Datenerhebung platziert werden. Entsprechend endete der Fragebogen mit den für die Befragten weniger fordernden und spannenden demographischen Fragen. Als Tool für die Umfrage wurde 'surveymonkey' verwendet.[43]

Grundsätzlich wurde beim Forschungsdesign darauf geachtet, die Konzentrationsfähigkeit der Teilnehmenden aufrecht zu erhalten. Für das Interview sowie für das Ausfüllen des Fragebogens wurden je zirka 30 Minuten eingeplant. Die Befragten hatten zudem Anspruch auf eine Pause zwischen den Teilen. Die geschätzten Zeiten wurden in den Pretests überprüft.

3.1.5. Wortwahl

Wie oben angeklungen sollten keine wertbesetze Begriffe in die qualitative oder quantitative Forschung einfliessen (Diekmann, 2007, S. 481). In den Interviews sowie im Fragebogen wurde aufgrund der Vermutung, dass die Befragten verschiedene Verständnisse und stark ausgeprägte Bewertungen für das Wort Selbstoptimierung haben, dieses bewusst nicht verwendet. Besonders wenn Selbstoptimierung negativ aufgefasst wird, könnten die Resultate in die jeweilige Richtung beeinflusst werden. In beiden Datenerhebungsverfahren wurde entsprechend der Begriff 'an sich Arbeiten' anstelle von Selbstoptimierung verwendet. Im Interview wurde der Begriff ohne eine vorgängige Definition verwendet, was den Studierenden die Möglichkeit gab, ihre spontanen Gedanken zu äussern und ihrem eigenen Verständnis zu folgen.

Im Fragebogen wurde der Begriff definiert, weil im Pretest in Erfahrung gebracht wurde, dass die Testinterviewten einerseits eine Definition hilfreich fänden und andererseits ohne Definition die selbstoptimierenden Tätigkeiten nicht genügend als solche verstanden haben. Im nächsten Kapitel wird das Testen der beiden Methoden sowie die daraus resultierenden Veränderungen ausgeführt.

3.2. Pretests

Ein Pretest eines Datenerhebungsverfahren ist notwendig, weil, wie Backstrom und Hursh (1963) betonten, kann keine noch so grosse intellektuelle Übung den Test einer Methode ersetzen, die für die Anwendung an Menschen entwickelt wurde. Ein Pretest stellt einen Probelauf des Datenerhebungsverfahrens, in dieser Arbeit des Interviews und anschliessendem Fragebogen, dar. Mit Hilfe dieser Probeläufe können Probleme im Datenerhebungsdesign, der Schulung des Personals sowie bei logistischen Abläufen erkannt und vor der finalen Datenerhebung behoben werden (UCLA CTSI Community Engagement & Research Program, n.d.).

Mit einer Zielprobanden*innenzahl von 30 Personen wurden sechs Personen zum Pretest eingeladen.[44] Je näher die Bedingungen der Pretests am geplanten Datenerhebungsdesign sind, desto wirkungsvoller sind sie. Die Bedingungen der Pretestsinterviews stimmten, abgesehen von inhaltlichen Veränderungen, mit denen der Interviews überein. Lediglich die Zielgruppe musste aufgrund der andauernden Prüfungsphase der Wirtschaftstudierenden auf Studierende aller Fächer ausgeweitet werden. Für die Pretests wurde die Debriefing Methode ausgewählt: Mit dieser wird die Datenerhebung in Echtzeit durchgeführt und danach in einem vorstrukturierten Gespräch die Datenerhebung reflektiert[45]. Den Testinterviewten wurden allgemeine Fragen zur Informationsmenge, Länge, Struktur, den Antwortmöglichkeiten sowie dem Fragenübergang gestellt. Zudem wurden die meisten Fragen erneut einzeln vorgelesen und die Schwierigkeit, das spezifische Verständnis der Frage oder von in der Frage vorhandenen Begriffen geprüft. Die Befragten wussten im Voraus, dass neben dem Inhalt ihrer Antworten vor allem ihr Verständnis der Fragen untersucht wird. Im Pretestprotokoll[46] sind alle Debriefing Fragen sowie die Einführung einsehbar. Während der Pretests beobachtete und protokollierte die Interviewerin zudem Signale wie starkes Zögern, Stirnrunzeln, Lachen oder andere Reaktionen und liess diese in die Auswertung der Pretests einfliessen.

Die Auswertung der Testinterviews und Fragebogen beschränkte sich darauf in Erfahrung zu bringen, ob die Fragen wie intendiert verstanden wurden. In einem Übersichtsdokument[47] wurden die Rückmeldungen zu jeder Interview- und Fragebogenfrage aller Befragten gesammelt. Wenn alle Personen die Frage korrekt verstanden haben, wurde keine Änderung vorgenommen. Wenn mehrere Personen Verbesserungsvorschläge hatten, Miss- oder- Unverständnis äusserten, wurde die Fragenformulierung verbessert, sofern von einer Verbesserung für alle Befragten ausgegangen werden konnte. Aufgrund des begrenzten Rahmens dieser Arbeit sind alle Veränderungen von Interviewleitfaden und Fragebogen im Drive hochgeladen und dort dokumentiert. Die groben Veränderungen seien hier erwähnt: Statt der Regelmässigkeit[48] als Indikator für die Intensität der Selb-

[43] Alle Fragen des finalen Fragebogens können unter diesem Link auf SWITCHDrive im Originaldesign eingesehen werden.

[44] Wie von Diekmann (2007, S. 485) beschrieben ist 20% der Zielmenge als Zahl der Testinterviewten nicht unüblich.

[45] Die Debriefing Methode hat gegenüber der Protokoll- und anderen Methoden den Vorteil, dass die Interviewerin durch die Durchführung des Datenerhebungsprozesses in Echtzeit ihre Interviewfähigkeiten verbessern kann und eine adäquate Einschätzung der Erhebungsdauer erhält. In anderen Pretestmethoden wird direkt nach der jeweiligen Interview-Frage eine Reflektionsfrage gestellt oder direkt bei der Frage eine Kommentarfunktion eingebaut, was den inhaltlichen Verlauf des Pretest zudem hätte beeinflussen können.

[46] Das Pretestinterviewprotokoll kann mit folgendem Link eingesehen werden.

[47] Mit diesem Link kann dieses Übersichtsdokument eingesehen werden. Aufgrund des begrenzten Rahmen dieser Arbeit wurde darauf verzichtet, ein Dokument mit allen Änderungen im Interviewleitfaden sowie im Fragebogen zu erstellen. Aus den Vergleichen der für die Pretests verwendeten Methodik und jener der Befragung werden alle Veränderungen sichtbar.

[48] Da Studierende aufgrund von langen Sommerpausen, intensiven Lern-

stoptimierungspraktiken wurde direkt Intensität verwendet. Darüber hinaus wurden gewisse Fragen zur besseren Übersicht zweigeteilt. Ansonsten wurden wie oben erwähnt zwei selbstreflexive Fragen hinzugefügt, um das Empfinden der Befragten während der Befragung abzuholen sowie einzelne Worte im Fragebogen durch besser verständliche Synonyme ersetzt. Das Pretesting diente zudem als Schulung für die Interviewerin, die sich einerseits in der neutralen Interviewführung und andererseits der Benutzung der Tools üben konnte.[49] Bei den Testinterviews bestätigten sich die Zeitaufwände: zwischen 20 bis 35 Minuten pro Interview und Fragebogen, durchschnittlich je 30. Entsprechend wurde fortan von einem durchschnittlichen Zeitaufwand von einer Stunde gesprochen.

3.3. Durchführung

3.3.1. Vorbereitung und Rekrutierung

Ziel der Arbeit war es, Aussagen über Studierende der Universität St. Gallen hinaus für die ganze deutsch-sprachige Schweiz treffen zu können[50]. Die Befragten wurden aus dem erweiterten Bekanntenkreis rekrutiert. Neben der Kontaktaufnahme mit Bekannten schrieb die Interviewerin in folgende Gruppenchats: WhatsApp-Gruppe mit allen Bachelorstudierenden der Volkswirtschaftslehre der Universität St. Gallen, WhatsApp-Gruppe mit allen Mitgliedern des 'Social Business Club' der Universität St. Gallen, Slack-Gruppe[51] mit allen Mitgliedern aller Nachhaltigkeitsorientierten Studierendenvereinen der Universität St. Gallen sowie Slack Gruppe mit allen oikos[52] Mitgliedern der Universität St. Gallen. Generell wurde entsprechend Diekmann's Empfehlungen (Diekmann, 2007, S. 444) kooperativ und motivierend kommuniziert. Die Befragten sollten wissen, dass ihrer Meinung Bedeutung zugemessen wird und dass die Befragung idealerweise sogar Spass macht. Die originale Rekrutierungsnachricht kann im Anhang eingesehen werden.

Die Teilnehmer*innen wussten lediglich, dass es in der Befragung um Selbstbilder geht und woher die Inspiration für das Thema stammte. Die interessierten Studierenden meldeten sich sodann per privater Nachricht via Slack oder WhatsApp bei der Interviewerin. Diese schlug den Interessierten sodann über ein aktuelles Foto des Interviewplans freie Termine vor. Alle Interviewplätze wurden in einem Zeitfenster von zwei Wochen angeboten, das zwei Wochen nach der Anfrage begann. Beginnend um 9:00 Uhr wurden im zwei Stunden Takt bis 20:00 Uhr Termine angeboten. Grüne Felder bedeuteten, dass die Termine noch frei sind. Eine rote Färbung des Terminfeldes bedeutete, dass der Termin bereits belegt ist. Gelb gefüllte Felder verlangten weitere Abklärungen. Sobald an einem Tag drei Interviews eingetragen waren, wurden keine Termine mehr an diesem Tag angeboten.[53] Nach zwei Interviews am Stück wurde der nächste Termin gelb markiert, sodass die Interviewerin zwischen dem zweiten und letzten Interview mindestens eine Stunde Pause machen konnte. Innerhalb von einer Woche waren entsprechend dieses Verfahrens bereits 20 der 30 Plätze gefüllt. Mit den ausstehenden Nachrichten konnten in der Woche darauf alle Plätze sowie ein Puffer am letzten Samstag gefüllt werden.

Die interessierten Studierenden erhielten bis zum Interview keine weiteren Informationen zum Inhalt des Interviews und Fragebogens. Sobald ein Termin vereinbart wurde, bedankte sich die Interviewerin im Voraus und teilte auf Nachfrage mit, dass keine Vorbereitung benötigt ist. Am Vortag des jeweiligen Interviews erhielten die Studierenden eine Erinnerung[54] an den Termin sowie den Link zum virtuellen Treffen auf Zoom.

3.3.2. Interview und Fragebogen

Die Interviewerin orientierte sich während jedem Interview am Interviewprotokoll[55], in welchem der Ablauf und die Verhaltensregeln für die Durchführung der Interviews und Fragebögen dokumentiert sind. In der gesamten Kommunikation vor dem Interview sowie der Begrüssung war das oberste Ziel, ein angenehmes Interviewklima zu schaffen. Dies ist insbesondere wichtig bei narrativen Interviews, weil die Befragten "aus Gründen der Anerkennung die Darstellung ihrer Biographie" (Mecke, 2001, S. 48) verändern könnten. Eine nicht aufgezeichnete und lockere Begrüssung[56] hatte zum Ziel, diesen Effekt zu begrenzen. In der Begrüssung wurden die Interviewten zudem um ihr Einverständnis zur Gesprächsaufzeichnung gefragt.

Nach dem Aufzeichnungsbeginn las die Interviewerin eine vorformulierte Einleitung vor, in welcher sie sich vorstellte und den Ablauf der Interviews sowie das anschliessende Ausfüllen des Fragebogens beschrieb. Es folgte eine deklarierte 'Warm-up' Frage, die die Befragten einerseits ins Erzählen bringen und andererseits langsam an das Thema heranführen sollte[57]. Anschliessend leitet die Interviewerin die

phasen und flexibler Stundenplanung eine relativ schwankende Alltagsgestaltung haben empfanden es die Befragten als schwierig, die Regelmässigkeit über ein Jahr zu bestimmen.

[49] Entsprechend wurden die Pretest Interviews wie die geplanten Interviews aufgezeichnet und transkribiert. Mit folgendem Link können die Transkripte und Audiodateien abgerufen werden.

[50] Da lediglich drei der Befragten nicht an der Universität St. Gallen (sondern an der Universität Basel, zum Teil mit Nebenfach an der Universität Luzern) studieren, sind die Aussagen in diesem Kontext anzusehen.

[51] Slack ist ein Kommunikationsprogramm, welches den Austausch von Nachrichten zwischen Individuen sowie innerhalb Gruppen ermöglicht und von vielen Studierendenvereinen der Universität St. Gallen verwendet wird.

[52] Oikos ist der grösste Nachhaltigkeitsverein der Universität St. Gallen, in welchem die Interviewerin zum Zeitpunkt der Anfrage Mitglied war. Die Interviewerin war zudem zwei Jahre im 'Social Business Club' aktiv.

[53] Aufgrund Terminverschiebungen aus der ersten Interviewwoche wurden an einem Tag in der zweiten Woche eine Ausnahme gemacht und vier Interviews durchgeführt. Unter diesem Link kann der Interviewplan eingesehen werden.

[54] Eine Beispielsnachricht kann im Anhang B eingesehen werden.

[55] Sowie dem darin enthaltenen finalen Interviewleitfaden.

[56] In welcher spontan eine Frage zum bisherigen Tag formuliert und darüber informiert wurde, dass der Hund der Interviewerin eventuell bellen könnte.

[57] Die Antworten auf die 'Warm-up' Frage wurden dokumentiert, fliessen jedoch nicht in die Auswertung ein.

Befragten durch alle Interviewfragen und versucht durch Wiederholen (von Teilen) der gegebenen Antworten und die aktive Verknüpfung mit den nächsten Fragen den Fluss des Interviews aufrecht zu halten und die Teilnehmer*innen beim Erzählen zu unterstützen. Hauptsächlich ist die Interviewerin bedacht, aktiv zuzuhören sowie die Erzähllinie zu verfolgen. Die Befragerin hält sich an den Interviewleitfaden und weicht nur davon ab, wenn Befragte massiv[58] anstehen. Wenn die Antworten der Befragten die gestellte Frage nicht beantworten, fragt die Befragerin nach. Dieses Nachfragen erfolgt beispielsweise "mittels Anknüpfung an bereits Gesagtes" oder Stimulation "zu weiteren detaillierten Erzählsträngen" (Mecke, 2001, S. 48) wie beispielsweise 'du hast angedeutet...'. Die Befragerin reagiert wertfrei auf die Antworten und nimmt diese neutral zur Kenntnis. Sie bedankt sich lediglich für das Beantworten der einzelnen Fragen. Bei weniger plausiblen oder unklar ausgeführten Aussagen sowie bei Unterbrechungen aufgrund von schlechter Internetverbindung oder Lärm fragt die Interviewerin nach, ob sie die bisherigen Aussagen richtig verstanden habe. Nachdem die Befragten alle Interviewfragen inklusive der Ergänzungswunschfrage beantwortet haben, frägt die Interviewerin die Befragten, ob sie eine fünfminütige Pause wünschen, um den Kopf zu durchlüften oder sich etwas zu trinken zu holen.

Nach einer fünfminütigen Pause oder direkt im Anschluss leitet die Interviewerin entsprechend dem Protokoll zum Fragebogen über. Sie erläutert das Vorgehen und erklärt, dass die Befragerin nach dem Ausfüllen des Fragebogens gerne inhaltliche Fragen zur Bachelorarbeit beantwortet. Die Befragerin wiederholt zudem, dass sich die Befragten Zeit lassen dürfen und sich melden können, wenn sie den Fragebogen abgesendet haben. Sie teilt dann den Link zum Fragebogen im Zoom-Meeting-Chat und stellt ihr Mikrofon auf stumm. Im Anschluss füllen die Befragten den Fragebogen selbstständig aus, während die Befragerin im Zoommeeting bleibt. Es steht den Befragten frei, ob sie sich stumm stellen und/oder ihre Kamera für das Ausfüllen des Fragebogens ausschalten wollen. Die Interviewerin behält ihre Kamera an, um den Befragten ihre Anwesenheit zu signalisieren. Die Befragerin liest oder schreibt, während die Befragten den Fragebogen ausfüllen, etwas auf ihrem Laptop. Dabei ist ihr Blick nicht auf die Uhr oder auf den*die Teilnehmer*in gerichtet, damit möglichst wenig Druck aufgebaut wird. Idealerweise wird durch die Anwesenheit der Befragerin Druck genommen, da starke Verständnisprobleme direkt angesprochen werden können. Zudem wird garantiert, dass zwischen dem Interview und dem Fragebogen nicht zu viel Zeit verstreicht, der Fragebogen alleine und vollständig beantwortet und abgeschickt wird. Nach dem Ausfüllen des Fragebogens fand mit allen, die Interesse und Zeit hatten, ein informeller Austausch statt. Die Befragten durften insbesondere inhaltliche Fragen zur Bachelorarbeit sowie zum Prozess stellen. In gewissen Anschlussgesprächen standen andere oder private Themen im Vordergrund.

[58] Unter massivem Anstehen wurde insbesondere langes Zögern, wahrgenommene Unsicherheit oder Nachfragen verstanden.

3.4. Auswertung

3.4.1. Interviews

Zur weiteren Auswertung wurde die Tonspur der Interviews mit einem iPad aufgezeichnet und anschliessend mit descript[59] anonymisiert transkribiert. Die Interviewerin erstellte sich zu Effizienz-Zwecken eine Word Vorlage[60] mit der Struktur jedes Transkripts; Angaben zur Interviewzeit, der Person sowie den vorformulierten Einleitungs-, Übergangs- und Abschlusstexten. Diese Vorlage wurde dann jeweils mit den exakt transkribierten Fragen und Antworten der Befragten gefüllt. Die Interviewerin wurde mit 'I', die Befragte Person mit 'B' abgekürzt. Zudem wurde das Verhalten der Personen sowie Zwischenfälle wie Internetprobleme, langes Zögern oder Lachen in [eckigen Klammern] dazugeschrieben. Wenn die Befragten Begriffe von anderen Sprachen verwendeten, wurden diese in einfachen 'Anführungszeichen' niedergeschrieben. Zur weiteren Analyse der Interviews wurden die Transkripte aller Interviews einzeln in die ATLAS.ti Software geladen. In dieser wurde die qualitative Inhaltanalyse inspiriert von Mayring gemacht[61].

Die Schritte der qualitativen Inhaltsanalyse nach Mayring lassen sich folgendermassen zusammenfassen: Zuerst wird die Forschungsfrage, das Material sowie die Entstehungssituation dieses Materials festgelegt.[62] Um eine systematische und zuverlässige Analyse qualitativer Daten, relevante Themen und Muster zu identifizieren und daraus Schlüsse zu ziehen, die zur Beantwortung der Forschungsfrage beitragen, wird anschliessend eine Analysetechnik sowie ein passender Ablauf definiert. Laut Mayring wird im Voraus definiert, wie gross die Kodiereinheit, Kontexteinheit und Auswertungseinheit sind.[63] Die strukturierte Form der Inhaltsanalyse nach Mayring wertet dann die Texte mithilfe von definierten Kategorien aus, die dann entsprechend des Ablaufs auf das Textmaterial angewandt werden. Anschliessend werden die Ergebnisse in einem Analyseprozess zusammengetragen und entsprechend der Forschungsfrage interpretiert (Mayring, 1991)[64].

[59] Descript ist eine "All-in-One-Videobearbeitungsplattform", die zudem Transkriptionsdienste anbietet und vielfach für die Bearbeitung von Podcasts eingesetzt wird.

[60] Die Vorlage kann unter folgendem Link eingesehen und heruntergeladen werden.

[61] Vgl. Kodierleitfaden im Anhang B sowie Kapitel zur qualitativen Inhaltsanalyse von Mayring in (Flick & Kardorff, 2006, S. 468-474)

[62] Vorgängige Schritte der Mayringschen Inhaltsanalyse wie die Festlegung der Forschungsrichtung, Fragestellung, Material sowie Entstehungssituation und formale Charakteristika dieses Materials wurden in vorherigen Methodenkapitel bereits beleuchtet und werden für dieses Unterkapitel als geklärt angenommen.

[63] Die Auswertungseinheit beschreibt, in welcher Reihenfolge die Texte nacheinander kodiert werden. Die Kodiereinheit legt den kleinstmöglichen Textbestandteil fest, der kodiert werden darf, während die Kontexteinheit den grössten beschreibt. Alle sind in dieser Arbeit im Kodierleitfaden beschrieben.

[64] Bereits beim Erstellen des qualitativen Forschungsdesigns wurde auf die nachfolgenden Gütekriterien qualitativer Forschung: Objektivität, Relia-

In dieser Arbeit wurden sowohl zuvor definierte Kategorien (deduktives Verfahren) als auch neu entstandene Kategorien verwendet (induktives Verfahren). Entsprechend wurde über mehre Schritte ein Kategoriensystem entwickelt, welches die relevanten Inhalte der Transkripte wiedergeben kann. Jede Kategorie, im folgenden Kode genannt, muss abgrenzend definiert werden. Dafür können Zitate aus der Theorie oder dem Textmaterial verwendet werden. Im Kodierleitfaden[65] wurde jede Kategorie beschrieben. Für jede Frage sind zudem die Kodierregeln, die Regeln zur Verwendung der Kodes, festgelegt. Das Setzen und Verändern von Kodes ist im Kodierleitfaden zur Nachvollziehbarkeit dokumentiert[66].

3.4.2. Fragebogen

Die Auswertung des Fragebogens diente als Kontrolle und Ergänzung der Interviewergebnisse. Dazu wurde zum einen die automatisch generierte Übersichtsauswertung von Surveymonkey verwendet[67] und zum anderen eigene Berechnungen mit dem Statistikprogramm R durchgeführt. Sozialdemographische Angaben wurden der Übersichtstabelle entnommen, während in R[68], inspiriert von der Auswertung des PERMA-Profilers und dem Moral Foundations Fragebogen, Summen und Durchschnittswerte[69], über vordefinierte Kategorien berechnet wurden. Bereits bei der Erstellung des Fragebogens wurde definiert, welche Aussagen zu welcher Kategorie gehören. Eine Übersicht kann dem Anhang B entnommen werden.

Für die Tätigkeits- und Gründeblöcke wurde ebenfalls mit der durchschnittlichen Zustimmung gearbeitet. Es wurde jeweils die mittlere Intensität der Arbeit an sich selbst pro Tätigkeitsbereich zusammengefasst um zu vergleichen, in welchen Bereichen die Befragten mehr oder weniger intensiv an sich arbeiten. Die Begründungen für die Arbeit an sich selbst wurden pro Bereich sowie insgesamt ausgewertet. Es wurde angenommen, dass stärkere Zustimmung zu Begründungen der Arbeit an sich selbst mit stärkerer Motivation, möglicherweise sogar Sinnvermögen einhergehen. Basierend auf den durchschnittlich beliebtesten Gründen pro Bereich sowie insgesamt sollten Rückschlüsse auf die Selbstrechtfertigungsnarrative ermöglicht werden. Zudem konnten die genannten Motivationen aus den Interviews überprüft werden. Die Antworten, die die Befragten frei zu den vorgeschriebenen Tätigkeiten und Gründen hinzuschreiben konnten, wurden individuell gelesen und ausgewertet.[70]

Neben der bisher beschriebenen Auswertungsmethodik bestand bei der quantitativen Auswertung eine gewisse Offenheit für die Untersuchung von Trends sowie der Anwendung anderer quantitativer Zusammenhänge. Die Antworten aus dem Fragebogen wurden zudem auf Korrelation geprüft. Es wurden die Antworten mit der höchsten Zustimmung pro Frage als Referenz analysiert und, wo basierend auf der Theorie Zusammenhänge möglich wären, wurden die Daten auf Unterschiede zwischen Personen verschiedener Geschlechtsidentitäten, Glaubensrichtungen sowie dem Besuchen des Kurses ‘Ökonomie des Glücks’ untersucht. Die Antworten auf die selbstreflexiven Fragen zum Empfinden während des Beantwortens wurden in die jeweiligen Interviewtranskripte kopiert und entsprechend des Kodierleitfadens in ATLAS.ti kodiert. R diente zudem als geeignetes Programm zur Visualisierung der Resultate. Der Kode zur Erstellung aller Darstellungen kann entsprechend im R-Skript[71] nachvollzogen werden. Es wurden vorab keine Kriterien zur Ausschliessbarkeit des Fragebogens definiert.[72]

Durch die im Anhang D aufgeführten Links kann auf die Rohdaten, die Interviewtranskripte sowie weiterte für die Arbeit zentrale Dokumente zugegriffen werden. Bevor auf die Resultate der Befragungen eingegangen wird, werden die Befragten im nächsten Kapitel anhand sozio-ökonomischer Kriterien beschrieben.

3.5. Befragte

17 der befragten 30 Personen identifiziert sich mit dem männlichen Geschlecht.[73] Die Mehrheit[74] der in dieser Bachelorarbeit befragten Personen war zum Zeitpunkt der Befragung zwischen 20 und 24 Jahren alt.[75] Mehr als zwei Drittel der Befragten befand sich im Bachelorstudium[76]. Die

bilität, Validität und Handlungsorientierung geachtet (Flick & Kardorff, 2006, S. 320 f.). Ebenso wurde auf Gütekriterien quantitativer Messungen (Diekmann, 2007, S. 247-260). Für weitere Ausführungen sei auf die hier aufgeführten Primärquellen verwiesen.

65 Der Kodierleitfaden kann im Anhang B eingesehen werden.

66 Aus ATLAS.ti wurde zudem ein Report aller angewendeten Kodes (Zitat-Manager) sowie der Veränderung der Kodes (Kode-Manager) zur Nachvollziehbarkeit exportiert. Zudem wurden direkt in den Transkripten in ATLAS.ti Tipp- und Rechtschreibefehler korrigiert.

67 Alle Auswertungen zum Fragebogen können unter folgendem Link eingesehen werden.

68 Die gesamte Aufbereitung und Auswertung der Daten kann im R Skript unter nachfolgendem Link eingesehen werden.

69 Im PERMA-Profiler werden die Aussagen zu jeder Kategorie aufsummiert. Basierend darauf findet die Auswertung statt. Der Moral Foundations Fragebogen arbeitet mit dem durchschnittlichen Zustimmungswert pro Kategorie und Person.

70 Die Antworten wurden entweder vorhandenen Tätigkeiten zugeordnet oder separat betrachtet. Es ging dabei darum, ein grundsätzliches Bild der ausgeführten Selbstoptimierungspraktiken zu erhalten.

71 Mit folgendem Link kann das R Skript heruntergeladen werden. Diesem Ordner können zudem die Rohdaten entnommen werden. Entsprechend der englischen Programmsprache wurde das ganze Skript in englischer Sprache verfasst.

72 Dies insbesondere, da angenommen wurde, dass die virtuelle Anwesenheit der Interviewerin ausreichend kontrollierend auf die Teilnehmenden wirkt. Im Nachhinein wäre es sinnvoll gewesen, Ausschlusskriterien wie beispielsweise eine Mindest- und Maximalzeit zu definieren, in welcher angenommen wird, dass der Fragebogen sinnvoll und konzentriert ausgefüllt wurde. Es kann dennoch gesagt werden, dass keine befragte Person weniger als zehn Minuten oder mehr als eine Stunde für die Beantwortung des Fragebogens benötigte. Zudem hätte mit einer Kontrollfrage die Aufmerksamkeit der Befragten überprüft werden können. Aufgrund der geringen Anzahl der befragten Personen wurde darauf verzichtet. Wie in den Limitationen erneut aufgegriffen, stellt das Einfügen einer Kontrollfrage einen Verbesserungsvorschlag für zukünftige Forschung dar.

73 Die restlichen 13 identifizieren sich als Frauen.

74 Es handelt sich dabei um 24 Personen.

75 Eine Person war zum Zeitpunkt der Befragung jünger als 20 Jahre alt, eine älter als 27. Die restlichen Befragten waren zwischen 24 und 27 Jahre alt.

76 Es handelt sich um 23 Personen. Darunter zwölf Personen, die Volkswirtschaft studieren, acht Internationale Beziehungen und drei Betrieb-

meisten Befragten sind im Studium bereits fortgeschritten – lediglich eine Person befindet sich im ersten Studienjahr. Neben dem Studium arbeiteten 14 Personen, mit unterschiedlichen Pensen, bezahlt. Neun interviewte Personen engagieren sich zudem freiwillig.

Mehr als die Hälfte der Personen[77] ordnet sich selbst der europäischen oder schweizerischen akademischen Mittelklasse zu. Neun Personen geben an, dass die erbbedingte Mittelklasse ihre sozio-ökonomische Situation am besten beschreibt. Es stammen entsprechend wenige Befragte aus einer kleinbürgerlichen- oder Arbeiterklasse.

Hinsichtlich ihrer politischen Haltung beschreiben sich 40% der Befragten als grünliberal. Fast ein weiteres Drittel der Befragten beschreibt die eigene politische Haltung als soziodemokratisch. Lediglich zwei Personen geben an, dass ihre politische Haltung am besten als konservativ beschrieben werden kann. 'Liberal' und 'grün' sind zwei weitere vertretene politische Haltungen. Die Mehrheit der Befragten[78] beschreibt sich als 'nicht religiös'. Fünf Personen beschreiben sich dem christlichen Glauben angehörig, drei dem Muslimischen, während sich eine Person als Agnostiker*in beschreibt. Die restlichen Personen[79] sind spirituell und glauben 'an etwas Grösseres.'

Acht Personen besuchten den von Prof. Kolmar an der Universität St. Gallen angebotenen Kurs 'Ökonomie des Glücks', der die Inspiration für diese Arbeit lieferte.

4. Resultate

Mit der vorgängig beschriebenen Kombination qualitativer und quantitativer[80] Forschung wurden diverse Zusammenhänge zwischen Selbstoptimierung, gelingendem Leben und Heldenreisen untersucht. In diesem Kapitel sind die signifikantesten Ergebnisse illustriert.

Es sei zudem angemerkt, dass den Grafiken aufgrund der kleinen Stichprobe (N=30) keine Standardabweichungen hinzugefügt wurden. Zur Verständlichkeit des Kapitels sind hier erneut die Antwortmöglichkeiten auf den Fragebogen und die entsprechenden numerischen Werte aufgeführt. Es sei zudem darauf hingewiesen, dass die zitierten Ausschnitte der Interviewtranskripte entsprechend der Absatznummerierung aus ATLAS.ti ausgewiesen werden.

Die Resultate des Fragebogens und der Interviews wurden für dieses Kapitel thematisch zusammengefasst. Zuerst wird das Gesellschaftsbild der Befragten beschrieben. Anschliessend werden deren Selbstoptimierungspraktiken- und -motive aufgezeigt. Schliesslich werden die Vorstellungen eines gelingenden Lebens und die Verbindungen zur Arbeit an sich selbst beleuchtet. Abschliessend wird illustriert, inwiefern die Erzählungen der Studierenden Heldenreisen-Strukturen aufweisen.

4.1. Gesellschaftsbild und Zugehörigkeit

Das Narrativ einer leistungsorientierten Gesellschaft zieht sich durch Interviews und Fragebogen. Mehrere Personen[81] beschreiben die Schweiz in den Interviews als sogenannte Leistungsgesellschaft. "Ich glaube die Schweiz ist eine Leistungsgesellschaft, die sehr stark auf Leistung und Materialistisches fokussiert ist und auch der persönliche Wert darüber definiert ist, was du machst, was du alles erreicht hast in deinem Leben und weniger, wer du als Person bist" (Transkript Person 17, Absatz 40). Meritokratisch, zielorientiert, arbeitswillig, leistungsmessend, ambitioniert und nach Erfolg strebend sind Worte, die die anderen der 12 Personen nennen, die den Kode "Q1 Leistungs-und zielorientiert" im Interview zugewiesen bekommen haben. Wie Abbildung 2 zeigt, zieht sich das Bild der Leistungsgesellschaft auch im Fragebogen durch.

Im Fragebogen wurde das Gesellschaftsbild und die Zugehörigkeit zur Gesellschaft über die drei Achsen Leistungserwarten, Sinn- und Entwicklungsmöglichkeiten erfragt. Die Befragten stimmen den drei Aussagen zur Leistungserwartung im Schnitt leicht bis ganz zu. Der Leistungserwartung in der Schweiz wird, zumindest von den Frauen, stärker zugestimmt als dem Empfinden von Sinn und Tiefe. Die Entwicklungsmöglichkeiten werden bei Frauen und Männern durchschnittlich weniger stark gesehen als die Leistungserwartung und das Sinnempfinden in der Gesellschaft, wobei Frauen weniger Entwicklungsmöglichkeiten sehen als Männer.

Die Antworten auf die separat ausgewerteten Kontrollfragen, die nicht in der 'Ich-Form' formuliert wurden, zeigen, dass die Befragten das gesellschaftliche Glück weniger optimistisch einschätzen als die eigenen Entwicklungsmöglichkeiten, das Empfinden von Sinn und die Leistungserwartung. Obwohl die Formulierungen der Kontrollfragen[82] dabei eine Rolle gespielt haben könnten, sprechen auch in den Interviews nur zwei Personen[83] davon, dass Schweizer*innen glücklich sind, beziehungsweise glückliche Leben führen. Keine Person spricht wiederum davon, dass Schweizer*innnen unglückliche Leben führen[84]. Die Befragten stimmen der Aussage, dass in der Schweiz nur rein

swirtschaft oder Wirtschaftswissenschaften. Die Personen im Master studieren unter anderem 'Business Innovation', 'Banking and Finance' oder Deutsche Philosophie mit Wirtschaft. Eine Person beantwortete die Frage nicht.

77 Es handelt sich dabei um 17 Personen.

78 Es handelt sich dabei um 18 Personen.

79 Es handelt sich dabei um sechs Personen.

80 Bei der quantitativen Analyse wurde teilweise über den ursprünglich angedachten Summen- und Mittelwertvergleich der Aussagen hinausgegangen. Viele Berechnungen basieren auf dem arithmetischen Mittel und nicht der Summe. Dies einerseits um mit Übersichtlicheren Werten zu arbeiten: Der numerische Wert des Durchschnitts kann direkt in einen Zustimmungswert umgerechnet werden. Zudem wurden kaum Korrelationen zwischen den Subfragen einer Kategorie festgestellt (die ursprünglich in einer Achse hätten zusammengezählt werden sollen). Zudem wurde ausgewertet, welche Subfragen die stärkste Zustimmung erhalten hat.

81 Die Antworten sind kodiert mit "Q1 Leistungs-/ zielorientiert".

82 "In unserer Gesellschaft ist es für Individuen leicht, gl"ucklich/gesund zu leben."

83 Die Antworten sind kodiert mit "Q1 Glücklich/ gesund".

84 Zwei der befragten Personen sind jedoch in psychologischer Betreuung oder tendieren dazu, weil sie mit mentalen Gesundheitsproblemen kämpfen.

Table 1: Übersicht 7-stufige Likert Skala Fragebogen (Quelle: eigene Darstellung)

Aussage Zustimmung	Aussage Intensität	Numerischer Wert
Stimme voll und ganz zu	Voll und ganz	7
Stimme zu	Ziemlich	6
Stimme leicht zu	Leicht	5
Neutrale Haltung	Neutrale Haltung	4
Stimme leicht nicht zu	Fast nicht	3
Stimme nicht zu	Nicht	2
Stimme überhaupt nicht zu	Überhaupt gar nicht	1
Ich weiss nicht	-	0

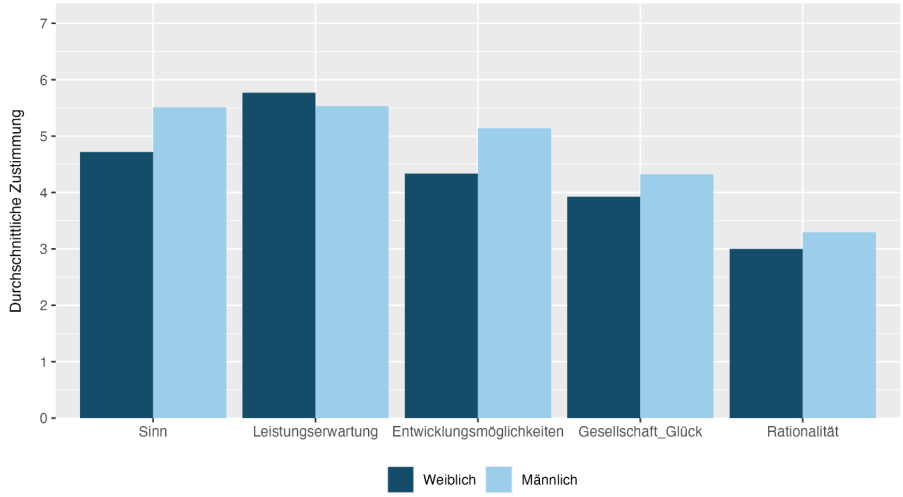


Figure 2: Wahrgenommenes Gesellschaftsbild (eigene Darstellung)

wissenschaftlich rationale Meinungen zählen, im Mittel leicht nicht zu.

Die Antwortmöglichkeiten[85], die die stärkste Zustimmung erhalten haben, sowie die Korrelationen innerhalb der Achsen, verstärken das bisherige Bild: 20-mal wird einer Aussage zur Leistungserwartung in der Gesellschaft am deutlichsten[86] zugestimmt, während die Wahrnehmung von Sinn in der Gesellschaft halb so oft und den Entwicklungsmöglichkeiten nur ¼ so oft so deutlich zugestimmt wird. Eine Zustimmung der Leistungserwartung korreliert zudem negativ mit der Zustimmung aller anderen Variabeln; dem Sinnempfinden, den Entwicklungsmöglichkeiten sowie dem generellen Glück und der Gesundheit der Bevölkerung.[87] Es sei hier auch erwähnt, dass fast die Hälfte der Befragten die Schweizer Gesellschaft in den Interviews als heterogen beschreiben.[88]

Die Fragen zum Gesellschaftsbild wurden insbesondere gestellt, um herauszufinden, inwiefern sich die Befragten in einer säkularen und/oder postheroischen Gesellschaft wahrnehmen und damit identifizieren können. Die Schweizer Gesellschaft wird von keiner befragten Person als heroisch oder postheroisch bezeichnet.[89] Gewisse der genannten Attribute lassen sich dennoch in heroisch und postheroisch einteilen, was nachkommend ausgeführt wird:

Vier Interviewte[90] nennen in die Schweizer Gesellschaft traditionell, wertetreu und konservativ. Konservativ wird von den Befragten vielfach mit Sicherheit und Risikoaversion in Verbindung gebracht. Die Interviewten beschreiben den Konservatismus auch im Zusammenhang mit langsamen, starren (politischen) Prozessen und der Engstirnigkeit der Schweizer Bevölkerung. Obwohl einzelne Personen[91] von starkem Gerechtigkeitssinn in der Schweiz sprechen, kann

[85] Damit sind alle Antwortmöglichkeiten der Frage zum Gesellschaftsbild gemeint.

[86] Oder gleich deutlich. Obwohl 30 Personen den Fragebogen ausgefüllt haben, sind 35 Antworten am deutlichsten zugestimmt worden, weil den Antworten gleich stark zugestimmt werden konnte und diese Antworten ebenfalls enthalten sind. Im R Skript sind die Berechnungen unter 'top answers' zu finden.

[87] Vgl. Korrelations-Tabelle Gesellschaftsbild

[88] Insgesamt 13 Personen. Die Antworten sind kodiert mit "Q1 Heterogen". Zum Teil zusammenhängend damit oder, weil die Personen nicht Schweizer*innen sind sowie aus anderen Gründen (wie die eigene Subjektivität) limitieren diese die Aussagekraft ihrer Antworten. Die Antworten sind kodiert mit "Q1 Limitationen".

[89] Dies kann daran liegen, dass die Personen keinen Bezug zu den Begriffen heroisch und postheroisch haben, worauf in der Diskussion im Kapitel 5.1 erneut eingegangen wird.

[90] Die Antworten sind kodiert mit "Q1 Konservativ/ engstirnig".

[91] Die Antworten sind kodiert mit "Q1 Verantwortungsbewusstsein/ Moral".

von den Antworten der Studierenden nicht auf eine breit verankerte moralische Ordnung in der Schweiz geschlossen werden, die das Zusammenleben durch gemeinsame Tugenden und Prinzipien leiten[92]. Keine Person nennt einen gemeinsamen Glauben, der eine solche Ordnung mit sich bringen würde.[93] Person 6 beschreibt die Schweizer Gesellschaft als eine, die "mittlerweile ein bisschen 'disconnected' von der eigenen Geschichte" ist und fühlt sich, obwohl die Person in der Schweiz aufgewachsen ist, "manchmal ein bisschen fremd und glaubt, dass das mit der Gesellschaft zusammenhängt" (Transkript, Absatz 38).[94]

Die demokratischen Strukturen werden von 1/3 der Befragten in der Frage nach dem Gesellschaftsbild angesprochen. Die Mehrheit der Befragten sieht die direkte Demokratie mit ihren Partizipationsmöglichkeiten positiv[95]. Andere äusserten sich ambivalenter zum demokratischen System der Schweiz; es könne herausfordernd sein oder in Populismus ausufern[96]. Person 23 beschreibt die Schweiz als "ein komplexes Produkt aus bisschen freierer Politik - eben direkter Demokratie – und trotzdem relativ strikten Strafen, falls doch irgendwas passieren sollte" (Transkript, Absatz 48), womit die Person ein weiteres postheroisches Merkmal, die geordnete Organisation der Gesellschaft anspricht. Person 10 bezieht sich auf das staatliche 'Management' der Gesellschaft: "Und ich bin super super dankbar für das politische System, das wir hier haben – wie der ganze Staat aufgebaut ist, die Institutionen wie das alles funktioniert. Staatlich sind wir schon eher effizient, würde ich sagen" (Transkript, Absatz 48). Neun Personen sprechen das Funktionieren, die Regeln oder Striktheit der Schweizerischen Gesellschaft an. Person 23 beschreibt regelrecht, wie Moral in den Schweizer Strukturen verankert ist:

> "Ich würde sagen, dass die Schweizer Gesellschaft sehr funktionell ist. Durch viele Regeln und das Eingreifen vom Staat ist es möglich, dass eine Gesellschaft aufgewachsen ist, die verantwortlich führt. Aber eben nicht nur durch Regeln, sondern durch Erziehung würde ich sagen, dass Kinder in Familien gross wachsen, die Verantwortlichkeit führen, füreinander gegeneinander, und dass es viele Normen gibt, die weit partizipiert sind, und darum finde ich ist es eine sehr gut funktionierenden Gesellschaft" (Transkript, Absatz 48).

Person 14 spricht von einer Rationalität "was den Schweizern sehr wichtig ist generell, ist, dass es stimmt, dass es irgendwie 'logisch von dir ist' (Transkript, Absatz 38), dass es irgendwo ein Rational gibt für Sachen. Wir brauchen irgendwie strukturiertes Denken, damit wir die Sachen verstehen, wir haben vielleicht ein weniger emotionales Verständnis." Zwei weitere Personen[97] beschreiben die Schweizer Gesellschaft als vernünftig. Wie oben beschrieben, geht aus dem Fragebogen hervor, dass nicht nur wissenschaftlich rationale Meinungen in der Schweiz Geltung haben.

Neun Personen[98] sprechen von der Bedeutung von Konsum, Konkurrenz, Wachstum und Erfolg in der Schweiz. Dass die kapitalistische Tauschlogik auch auf die eigene Person bezogen wird, wird beispielsweise in folgendem Zitat von Person 18 deutlich: "Irgendwie habe ich schon das Gefühl, dass wir diese Einstellung haben, dass alles besser sein muss und man überall noch mehr (machen) und dass man sich wenig zufrieden gibt. Dass man nicht sagt, so jetzt habe ich genug, sondern, dass es immer weiter gehen muss" (Transkript, Absatz 48).

Adaptivität und Innovation werden in der breiten Frage zum Gesellschaftsbild von wenigen Studierenden[99] direkt angesprochen. Auf die Folgefrage antwortet eine Mehrheit der Studierenden, dass sie gesellschaftliche Erwartungen wahrnehmen, ein Leben lang an sich zu arbeiten. Folgende Antwort[100] deutet auf einen schnellen sozialen Wandel bis Anpassungsdruck hin:

> "Ja, ich glaube es wird immer stärker. Ich meine auch, wenn wir wenn den generellen Wachstumszwang anschauen von Unternehmen egal ob klein, oder so Startups, die vielleicht so schnell wie nur möglich schwarze Zahlen schreiben müssen oder grosse Firmen, die immer noch weiter wachsen müssen, obwohl sie eigentlich ihren Mehrwert schon erfüllen (ohne Wachstum), also von dem her ich glaube man versucht generell die Grenzen auszuloten und dadurch würde ich schon sagen, dass es erwartet wird, dass man immer an sich arbeiten muss, dass man nicht stehen bleiben soll – vor allem in der Schweiz, weil es wirklich immer schneller wird" (Transkript, Person 1, Absatz 43).

In der Interviewfrage zum eigenen Platz[101] antworten 1/3 der Personen, dass sie sich als Treiber von Innovation

[92] Lediglich positive Eigenschaften wie Zuverlässigkeit, Pünktlichkeit, Höflichkeit und Pflichtbewusstsein werden von 8 Personen beschrieben. Die Antworten sind kodiert mit "Q1 Zuverlässig".

[93] Aufgrund dessen kann von einem säkularen Gesellschaftsbild ausgegangen werden.

[94] Eine weitere Person (Person 19) nimmt die Schweizer*innen als "sehr auf sich bedacht, sehr individualistisch" wahr. Die Antwort ist kodiert mit "Q1 Individualisierend/ disconnected". Obwohl die Schweiz von drei Personen als kulturell offen bezeichnet wird, überwiegen die Personen, die die Schweizer*innen als zurückhaltend und distanziert wahrnehmen. Die Antworten sind kodiert mit "Q1 Kulturell offen & vernetzt" sowie "Q1 Zurückhaltend/ distanziert".

[95] Die Personen sprachen beispielweise von guten Kompromissen. Die Antworten sind kodiert mit "Q1 Demokratisch".

[96] Die Antworten sind kodiert mit "Q1 Demokratisch". Person 1 spricht von sozialer Ungleichheit im politischen System: Die die bereits Macht oder ein gutes Netzwerk haben, versuchen dies zu erhalten.

[97] Die Antworten sind kodiert mit "Q1 Pragmatisch/ rational".

[98] Die Antworten sind kodiert mit "Q1 Kapitalistisch/ kompetitiv".

[99] 4 Personen sprechen von Innovation, Forschung und dem guten Bildungssystem der Schweiz. Die Antworten sind kodiert mit "Q1 Adaptiv/ innovativ".

[100] Person 4 drückt sich im Absatz 44 des Interviewtranskripts ähnlich aus.

[101] Für die Beantwortung dieser Forschungsfrage weniger relevant oder häufig verwendete Kodes waren die Kodes "Q1: Entfaltungsmöglichkeiten", "Q1: Dankbar (eigene Person)" sowie "Q1: kein Teil davon".

und/oder Verantwortung sehen[102]. Innovation scheint unter anderem für Person 16 eng mit der Identität verknüpft zu sein "bin ich schon so der 'challenger'. Ich glaube ich rüttle gerne das Gewohnte auf mit Themen, die mir wichtig sind – also ein bisschen aktivistisch" (Transkript, Absatz 39). 2/3 der befragten Personen fühlt sich der Schweizer Gesellschaft zugehörig. Dieses Gefühl der Zugehörigkeit geht entweder auf gemeinsame kulturelle Aspekte oder das direkte private Umfeld zurück. Von gewissen Personen wird Zugehörigkeit funktional definiert; wie folgende Aussagen zeigen: "Ja, also ich fühle mich sehr wohl. Ich fühle mich nicht irgendwie abgehängt von der Gesellschaft. Ich bin ein sehr gut funktionierendes Teil dieser Gesellschaft. Ich habe Ambitionen und alles, was man so dafür braucht" (Transkript, Person 15, Absatz 40). Oder wie Person 13 es ausdrückt: "ein verlässlicher Teil des Netzwerks" zu sein (Transkript, Absatz 40). Person 17 spricht mehr von einer Anpassung an das System und davon sich gut angepasst zu haben (Transkript, Absatz 44).

Zehn Personen fühlen sich privilegiert in der Schweiz, ihrer Familie aufgewachsen zu sein oder in der Schweiz studieren zu dürfen. Ein kleinerer Teil der Befragten fühlt sich unsicher hinsichtlich ihres Platzes in der Gesellschaft[103]. Person 6 verwendet für dieses Gefühl folgende Metapher: "Ich brauche dieses Bild von einem Bahnhof gern, wo alle wissen, in welchem Zug sie steigen müssen und der Zug fährt sie einfach dort hin, sie müssen nur einsteigen und sitzen bleiben. Und ich habe einfach keinen Zug, so fühlt sich das ein bisschen an" (Transkript, Absatz 48).

Bei der Überleitung vom Gesellschaftsbild zu Selbstoptimierungspraktiken der Interviewten wurden diese gefragt, ob sie eine gesellschaftliche Erwartung wahrnehmen, ein Leben lang an sich arbeiten zu müssen, sowie in der Folgefrage, ob sie dies von sich selbst erwarten. Bei den gesellschaftlichen Erwartungen gehen die Meinungen der Studierenden auseinander. Acht Personen sprechen sich klar für[104], vier zögerlich für eine solche Erwartungshaltung aus, wobei nicht alle diese Erwartung als positiv beschreiben. Drei Personen vertreten eine klare und zwei eine zögerliche Gegenposition, dass keine solche Erwartungshaltung existiert. Die restlichen Befragten bekamen den Kode "kommt darauf an" zugeordnet, da sie die gesellschaftliche Erwartungshaltung des an sich Arbeitens von Faktoren wie den Entwicklungsmöglichkeiten, dem Umfeld der Personen sowie den Individuen selbst abhängig machen. Einige der Befragten sprachen in der Antwort auf diese Frage einen Unterschied zwischen dem universitären und dem gesamtgesellschaftlichen Umfeld an.[105]

In den Antworten zu den gesellschaftlichen Erwartungen konnten zwei Trends ausgemacht werden, die wiederum das bisherige Gesellschaftsbild verstärken: Ein Drittel der Befragten[106] spricht von einer Notwendigkeit, ein Leben lang zu lernen, um mit dem Wandel der Zeit mithalten zu können und das Funktionieren der Gesellschaft zu garantieren. Person 25 stellt fest "dass in der Schweiz so dieses Streben nach Wissen und Streben nach materiellem Erfolg und Status sehr stark verankert ist und man davon ausgeht, dass um ja (du hast vorhin 'Flow' gesagt) in diesem 'Flow' zu bleiben, dieses lebenslange Lernen vorausgesetzt wird" (Transkript, Absatz 50). Die Zustimmung von Person 26 "weil sonst das System nicht funktionieren würde. [...] Ein Leben langes nicht, aber einfach bis man pensioniert ist. Ohne diese Leistung wäre es sehr schwierig, dass irgendetwas funktionieren würde" (Transkript, Absatz 46), illustriert, dass in ähnlicher, funktioneller Form auch vom beruflichen Arbeiten gesprochen wird[107].

Im Gegensatz zur Einschätzung der gesellschaftlichen Arbeitserwartung fielen die Antworten auf die Frage nach den persönlichen Erwartungen weniger divers aus: 23 Personen erwarten klar von sich, dass sie an sich arbeiten. Drei Personen arbeiten je nach Definition des 'an sich Arbeitens' an sich, während eine Person klar nicht an sich arbeitet und drei zögerlich nicht.[108]

Abschliessend lässt sich sagen, dass sich durch die zahlreiche Nennung postheroischer Merkmale wie Adaptivität und Innovation, Konsum, Konkurrenz, Rationalität, Organisation, Demokratie und das Fehlen eines gemeinsamen Wertehorizont auf ein postheroisches Gesellschaftsbild mit hohen Leistungserwartungen schliessen lässt. Zudem erwarten die allermeisten Befragten von sich, an sich zu arbeiten. Das nächste Kapitel gibt Auskunft darüber in welchen Bereichen die befragten Studierenden mit welchen Methoden an sich arbeiten.

4.2. Selbstoptimierungspraktiken

Im Fragebogen, in welchem das an sich Arbeiten definiert wurde, gaben die Befragten in fünf vordefinierten Bereichen an, wie intensiv sie[109] an sich arbeiten.

Die Abbildung 3 zeigt die durchschnittliche Intensität, mit welchen die Befragten den aufgeführten Tätigkeitsbeispielen[110] in jedem Bereich nachgehen. Mit einem Blick auf die y-Achse zeigt sich ein zurückhaltendes Bild: Es wird in allen

[102] Die Antworten sind kodiert mit "Q1: Treiber von Innovation / Verantwortung (eigene Person)".

[103] Die Antworten sind kodiert mit "Q1: Unsicherheit / Suchen (eigene Person)".

[104] Die Antworten sind kodiert mit "Q2 Klares ja" sowie "Q2 Zögerliches ja", "Q2 Klares nein" und "Q2 Zögerliches nein".

[105] Den auf die Universität bezogenen Antworten wurde im Anhang C ein Kapitel gewidmet, da diese die aktuelle Lebenswelt der Studierenden prägt, aber für die Beantwortung der Forschungsfrage nicht relevant genug erscheint, um im Hauptteil der Arbeit beleuchtet zu werden.

[106] Die Antworten sind kodiert als "Q2/Q3 Leben langes Lernen".

[107] Insgesamt sprechen 8 Personen besonders vom beruflichen Arbeiten. Die Antworten sind kodiert unter "Q2 Ja, besonders beruflich".

[108] Es sei hier darauf aufmerksam gemacht, dass sich in den Antworten der Studierenden teils Unterschiede in den Verständnissen vom an sich Arbeiten abzeichnen, auf welche im nächsten Kapitel eingegangen wird.

[109] Die Befragten wählten auf einer 7-stufigen Likert-Skala aus, wie intensiv sie diverse Tätigkeiten im jeweiligen Bereich ausführen.

[110] Pro Bereich wurden zwischen sechs bis neun Tätigkeitsbeispiele aufgeführt, was die Resultate in die eine oder andere Richtung beeinflusst haben könnte. Die Studierenden hatten in jedem Bereich die Möglichkeit, eigene Tätigkeiten und die jeweilige Durchführungsintensität als Kommentar hinzuzufügen.

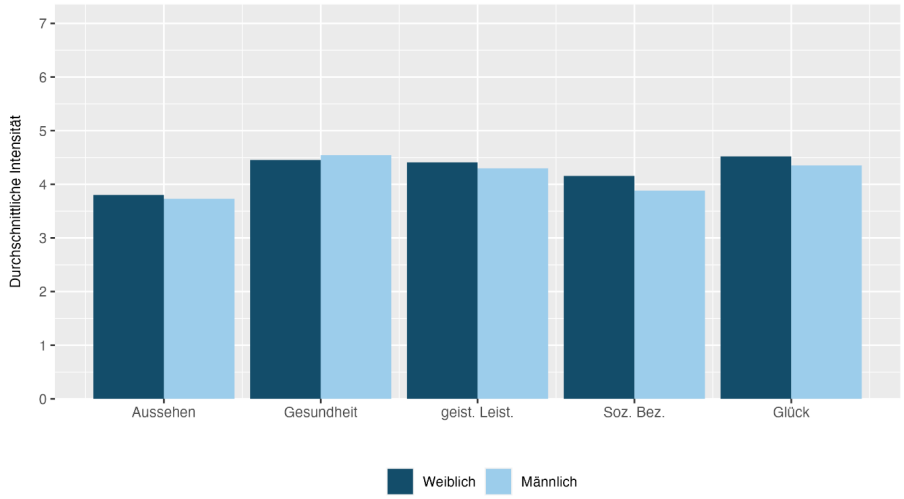


Figure 3: Durchschnittliche Intensität der Arbeit an sich selbst in verschiedenen Bereichen (eigene Darstellung)

Bereichen durchschnittlich zwischen fast nicht intensiv[111] bis leicht intensiv[112] an sich gearbeitet. Frauen arbeiten in allen Bereichen ausser der Gesundheit im Mittel intensiver an sich, wobei der Unterschied in den sozialen Beziehungen und dem Glück am grössten, aber nicht bedeutend gross ausfällt.

Entsprechend dem Fragebogen werden folgende Tätigkeiten von den Befragten in den jeweiligen Bereichen am intensivsten ausgeführt[113]: Modebewusst kleiden, stylen sowie gezielte Sportübungen (im Mittel leicht intensiv) im Bereich Aussehen[114]. Lernen, sich Wissen aneignen sowie gezielte sportliche Aktivität (im Mittel intensiv sowie leicht intensiv) für die geistige Leistungsfähigkeit[115]. Gesunde Ernährung sowie sportliche Aktivität (im Mittel etwas mehr als leicht intensiv) für die Gesundheit[116]. Für die Arbeit an den eigenen sozialen Fähigkeiten und sozialen Beziehungen[117] werden am intensivsten Meditieren und andere Achtsamkeitsübungen sowie Übungen zur Empathiefähigkeit durchgeführt (im Mittel zwischen neutral und leicht intensiv). Im Bereich Glück[118] wird am intensivsten Sport betrieben sowie eine Work-Life Balance hergestellt (im Mittel etwas über leicht intensiv).

Mit Blick auf die starken Erwartungshaltungen der Interviewten an sich selbst zu arbeiten, ist es plausibel anzunehmen, dass die Tätigkeitsbeispiele des Fragebogens nicht vollständig auf die Personen passen.[119] Mit den Antworten aus den Interviews wird ein holistischeres Bild der Selbstoptimierungstätigkeiten möglich, welches im Folgenden dargestellt wird:

Auf die Frage, in welchen Bereichen und wie die Befragten an sich arbeiten, antwortet Person 23: ”Ich würde das in persönliche und professionelle Entwicklung splitten” (Transkript, Absatz 68). Sechs weitere Personen[120] trennen die Arbeit an sich selbst in einen privaten und einen beruflichen Strang. Je 19 Personen sprechen in den Interviews vom an sich Arbeiten im professionellen sowie akademischen Bereich[121]. Während gewisse Personen wie Person 8 die Bereiche verbinden: das ist auch ein relevantes extra curriculares Engagement, das ich aktiv betreibe, um einerseits selbst mehr

[111] Übereinstimmend mit Wert 3.

[112] Übereinstimmend mit Wert 5. Während Wert 4 die neutrale Haltung abbildet.

[113] Um ein umfänglicheres Bild der Selbstoptimierungspraktiken zu erhalten, sind in den Fussnoten zudem relevante Tätigkeiten notiert, die die Befragten in Form von Kommentaren hinzugefügt haben. Relevant bedeutet in diesem Zusammenhang, dass im Fragebogen keine vergleichbare vorstandardisierte Tätigkeit enthalten war.

[114] Die Befragten nannten keine weiteren relevanten Praktiken im Bereich des Aussehens in der im Fragebogen dafür vorgesehenen Kategorie.

[115] Reflektieren und sich mit anderen Austauschen wurden je einmal als andere Praktiken genannt.

[116] Die Befragten führten hinsichtlich der Gesundheit bereits genannte Beispiele in der Funktion ‘Andere’ in eigenen Worten wie Psychotherapie, genügend, regelmässig und ausgewogen Essen aus. Neben der einmaligen Nennung von ‘Biohacks’ wurden keine relevanten Tätigkeiten genannt.

[117] Der Arbeit an sich selbst in sozialen Beziehungen und hinsichtlich sozialer Fähigkeiten fügten die Befragten Selbstsicherheit, eine optimistischere Lebenseinstellung erhalten als Selbstoptimierungstätigkeit hinzu und nannten wiederholt die Arbeit an der Kommunikationsfähigkeit sowie regelmässiges Kommunizieren.

[118] Im Bereich Glück äussert sich eine Person kritisch ”an meinem Glück arbeiten?” im offenen Tätigkeitsfeld. Andere fügten die Änderung der Lebenseinstellung, Zeit mit nahen Personen verbringen, Anwendung des Stoizismus auf das eigene Leben sowie die Priorisierung von freudigen Tätigkeiten (über solche, die weniger Spass machen).

[119] Möglicherweise war der Aufwand, im Fragebogen zusätzliche Tätigkeiten hinzuzufügen für die Studierenden zu hoch. Aus den Rückmeldungen des offenen Tätigkeitsfeldes ‘Andere’ konnten basierend auf maximal fünf Rückmeldungen pro Bereich keine Trends erkannt werden. Entsprechend wurden den erkennbaren Mustern in den Interviews besondere Beachtung geschenkt.

[120] Die Antworten sind kodiert mit ”Aufteilung Privat vs. Professionell”.

[121] Die Antworten sind jeweils kodiert mit ”Q4 Bereich Akademisch” und ”Q4 Bereich Beruflich”. Einige Befragten verbinden die Bereiche mit Aussagen wie ”akademisch/beruflich, hinsichtlich der Bildung/Karriere oder professionell/akademisch”, andere trennen die Bereiche. Bei 10 Personen sind beide Kodes vorhanden.

dazuzulernen und andererseits quasi den ‘potential employers’ zu zeigen: ”Okay, ich mach das, ich mach das, ich bin engagiert. Dann will man natürlich gute Noten schreiben, was auch ein ‘signaling effect’ ist, für den Beruf” (Transkript, Absatz 88), sehen andere Personen wie die Person 1 (Transkript, Absatz 51) die Ausbildung als etwas, durch das die Person ‘durch muss’ um den von ihr gewünschten Beruf ausüben zu können. Es lässt sich feststellen, dass das an sich Arbeiten im beruflichen Bereich der Studierenden stark auf das Erlernen von diversen Fähigkeiten und Suchen nach Erfahrung ausgelegt ist. Person 6 beginnt ihre Antwort beispielsweise wie folgt: ”Also ich glaube so, was am offensichtlichsten ist, ist alles, was Karriere anbelangt, sich da irgendwie ja wirklich zu einem ‘asset’ zu machen, sozusagen. Dass man nicht einfach nur ein Diplom in der Hand hat, sondern wirklich Erfahrung mitbringt, coole ‘insights’ hat zu Speziellem was auch immer” (Transkript, Absatz 62).

Übereinstimmend mit dem beliebtesten Bereich des an sich Arbeitens wurde im Interview die Methode sich Wissen und Fähigkeiten aneignen am häufigsten genannt. Beispiele sind jede Woche ein neues Kodierprojekt durchzuführen (Transkript, Person 23, Absatz 68) oder sich beim Lesen von Büchern Notizen zu machen ”um möglichst viel herauszunehmen” (Transkript, Person 20, Absatz 63). Nur in einem Drittel der Fälle wurde sich Wissen und Fähigkeiten aneignen direkt mit dem akademischen Bereich in Verbindung gebracht[122].

In den Interviews wurde deutlich häufiger von der Arbeit an der eigenen Person und/oder dem eigenen Glück[123] gesprochen als der Arbeit im Bereich Gesundheit[124]. Der Kode ”Q4 Bereich eigene Person / Glück” fand bei Heldenreisen-ähnlichen ‘Arbeiten’ wie dem Kampf gegen die inneren Ängste und Dämonen wie beispielsweise tiefem Selbstwertgefühl und Selbstzweifel Anwendung. Mehrfach wurde die Arbeit am eigenen Glück und/oder der eigenen Person mit der Methode Selbstreflexion genannt.

Selbstreflexion[125] sowie die Reflexion zusammen mit anderen Personen wurde insgesamt von 14 Personen als Methode des an sich Arbeitens genannt. Wie Person 12 das beschreibt, reflektieren die Befragten in verschiedenen Aspekten des Lebens ”Ich mach mir eigentlich über sehr viele Aspekte meines Lebens und meiner zwischenmenschlichen Kontakte, die ich habe, sehr viel Gedanken, zum Beispiel habe ich was Falsches gemacht, wie kann ich das ändern, wie habe ich jetzt irgendwie etwas überschritten, ‘ne Grenze oder so, und ich muss schon sagen ich, ich geb’ mir Mühe, auch wenn’s nicht immer funktioniert” (Transkript, Absatz 52). Inhaltliche Parallelen der Selbstreflexion bestanden insbesondere hinsichtlich des Umgangs mit Mitmenschen sowie der Verwendung des Tagebuchs oder den Austausch mit anderen Personen als Hilfsmittel für die Selbstreflexion. ”Und im persönlichen Bereich versuche ich oft zu hinterfragen, welche Dinge ich falsch mache, wenn ich mit anderen Leuten umgehe und versuche, die besser zu machen und da spreche ich auch manchmal mit Leuten, die mir wichtig sind, drüber und versuche daran zu arbeiten” (Transkript, Person 23, Absatz 68). Fünf Personen[126] nennen auf selbstreflexive Weise auch, dass sie gerne mehr oder weniger an sich arbeiten möchten.

Fünf Personen[127] nennen sich kennenlernen als Methode um an sich zu arbeiten. Unter sich kennenlernen wird unter anderem herauszufinden, was einen glücklich macht, das eigene Verhalten zu verstehen und die eigenen Stärken zu kennen, verstanden.

In Übereinstimmung mit dem vielfach genannten sozialen Bereich, nennen acht Personen[128] Austausch, Beziehungspflege und die Verbesserung der eigenen sozialen Fähigkeiten als Methode für die Arbeit an sich selbst. ”Familie und Freunde ist halt das bewusste Zeit nehmen und die Beziehungen pflegen” (Transkript, Person 10, Absatz 59). Der soziale Bereich wurde nach dem akademischen und beruflichen Bereich am häufigsten[129] in den Interviews angesprochen. Empathisch sein, die eigenen Emotionen kontrollieren und so möglichst konfliktfrei und harmonisch zu leben scheint den Studierenden in dem Bereich ein Anliegen zu sein. Auch das Aufbauen eines guten Netzwerkes wird von zwei Personen[130] als Arbeit im beruflichen Bereich gesehen.

Übereinstimmend mit der geringen Intensität der Arbeit an sich selbst im Bereich Aussehen im Fragebogen, wurde das Aussehen in den Interviews gar nicht genannt und sogar explizit von Sport und Gesundheit abgetrennt. Wenige Befragte[131] antworten, dass sie vollumfänglich an sich arbeiten, wenn auch die meisten Interviewten mehreren Bereiche nennen, in denen sie an sich arbeiten.[132]

Zusammenfassend kann gesagt werden, dass die Resultate des Fragebogens zwar leicht in der angegebenen

[122] Dies wurde durch die ATLAS.ti interne Kookkurrenz-Analyse von Kodes in Erfahrung gebracht. Zwei Kodes sind kookurrent, wenn sich die Zitate zu den Kodes überschneiden oder identisch sind.

[123] Genau die Hälfte der Personen spricht von der Arbeit an der eigenen Person und dem eigenen Glück. Die Antworten sind kodiert mit ”Q4 Bereich Glück / eigene Person”.

[124] Fünf Personen sprechen von der Arbeit an sich selbst im Bereich Gesundheit. Die Antworten sind kodiert mit ”Q4 Bereich Gesundheit”. Dass im Fragebogen die Arbeit an der eigenen Gesundheit im Vergleich zum Interview relativ hoch gewertet wurde, kann daran liegen, dass die im Bereich Gesundheit genannten Aktivitäten wie gesunde Ernährung und Sport im Interview separat als Methode und Sport separat als Bereich kodiert wurden. Diese wurden von fünf bis acht Personen genannt.

[125] Die Antworten sind kodiert mit ”Q4 Methode Selbstreflexion (sowie mit Anderen)”.

[126] Die Antworten sind kodiert mit ”Vergangenheit”.

[127] Die Antworten sind kodiert mit ”Q4 Methode sich selbst kennenlernen”.

[128] Die Antworten sind kodiert mit ”Q4 Methode Austausch, Beziehungspflege, soziale Fähigkeiten”.

[129] Es handelt sich dabei um 16 Personen. Die Antworten sind kodiert mit ”Q4 Bereich Soziales”.

[130] Die Antworten sind kodiert mit ”Q4 Netzwerk”.

[131] Die Antworten sind kodiert mit ”Q4 Vollumfänglich”.

[132] Basierend auf den Antworten des Fragebogens wurde untersucht, wie unterschiedlich intensiv die Befragten über alle Bereiche an sich arbeiten: Durchschnittlich arbeitet die Person, die über alle Bereiche am intensivsten an sich arbeitet zirka doppelt so intensiv an sich; wie die Person, die am wenigsten intensiv an sich arbeitet. Dies wird unter anderem in Abbildung 7 im Kapitel 4.5 illustriert.

Intensität[133], nicht aber in den genannten Bereichen und Methoden von jenen der Interviews abweichen. Die Befragten arbeiten hinsichtlich des Berufslebens, dem aktuellen akademischem Leben sowie im sozialen Bereich und jenem der eigenen Person an sich. Beliebte Methoden für die Arbeit an sich selbst stellen Wissensaneignung, Selbstreflexion und Beziehungspflege dar. Sich selbst kennenzulernen wird weniger häufig als Methode zur Arbeit an sich selbst genannt.

An dieser Stelle scheint es wichtig, die Diversität der Verständnisse des 'an sich Arbeitens' aufzuzeigen. Die meisten Personen beschreiben an sich arbeiten als aktiven Prozess. Einige sprechen in ihren Antworten von dazu lernen, ein Leben lang zu lernen oder sich weiterzuentwickeln, ohne die Begriffe weiter auszuführen.[134] Vielfach werden die Begriffe auf das Kennenlernen diverser Perspektiven im beruflichen oder akademischen Bereich bezogen. Person 26 beschreibt sich weiterentwickeln als automatischen "natürlichen Prozess", in dem sich jede Person weiterentwickelt, egal ob das erwartet wird oder nicht (Transkript, Absatz 50), während andere Personen Weiterentwickeln stark an systematische Selbstreflexion anlehnen oder das Weiterentwickeln darauf basieren. Bei wenigen Personen wird sich weiterentwickeln sogar mit sich kennenlernen, den eigenen Platz finden in Verbindung gebracht, wie Person 1 in folgendem Zitat beschreibt "wie navigier ich durch dieses System, Priorisierungen kennen, auch eigene Stärken (in dem Feld)" (Transkript, Absatz 51). Obschon das an sich Arbeiten, anhand der Reaktionen der Befragten, kaum Bestandteil des alltäglichen Vokabulars ist, gab es keine Anzeichen dafür, dass der Begriff nicht wie intendiert verstanden wurde.[135]

Manche Personen brachten den Begriff an sich Arbeiten in ihren Antworten direkt mit Selbstverbesserung oder Optimierung in Verbindung. Person 27 sieht in der Selbstverbesserung 'intrinsischen Wert' und sieht sie als Verpflichtung an: "Manche glauben an Gott, manche an die Evolution, aber alleine, dass wir hier sind, ist so absurd und jetzt sind wir in der Situation und jetzt müssen wir das Beste daraus machen und unsere Potenziale erreichen. Ich glaube, uns wurde damit eine 'Verpflichtung' mitgegeben (in Anführungszeichen)" (Transkript, Absatz 85). Während Person 28 "ein gewisses Aggressionsverhältnis" hinsichtlich Selbstoptimierung sieht: "Wenn beispielsweise im Gym immer mehr gedrückt werden, immer mehr verdient werden oder irgendeine andere Zahl vergrössert werden muss. Die Person ordnet danach ein, dass das nicht immer positiv sein muss, aber der Habitus wohl schon zu wachsen ist" (Transkript, Absatz 49). Person 6 formuliert vielmehr einen "Druck an der Uni und das Konstante 'sich vergleichen'. Es fühlt sich an als würde ich mich die ganze Zeit optimieren wollen. Ich glaube, da muss ich eher ein bisschen herunterschrauben und vielleicht ist an mir arbeiten dann halt einfach etwas anderes: zum Beispiel lernen im Moment zu sein" (Transkript, Absatz 62). Mehrere Personen erklären, dass sich beim an sich Arbeiten der eigene Fokus verschiebt, oder sie sogar davon wegkommen wollen.[136] Zusammenfassend geht aus den Interviews hervor, dass fast alle Personen sich aktiv weiterentwickeln wollen und dass dies wie beispielsweise von Person 27 positiv bewertet wird: "Ich erwarte von meinem Umfeld, ob's Freunde oder Familie sind, dass sie sich ständig weiterentwickeln, und ich glaube, das gleiche wird auch von mir erwartet. Aber das ist für mich eine sehr gute Sache" (Transkript, Absatz 55). Das nächste Kapitel widmet sich ausschliesslich den für die Selbstoptimierung genannten Gründen.

4.3. Motivation zur Selbstoptimierung

Die Abbildung 4 illustriert, dass über alle Bereiche die beliebtesten Gründe zur Selbstoptimierung das persönliche Glück, das kurzfristige Wohlbefinden, die Erreichung des eigenen Lebensziels sowie sich selbst kennenzulernen sind. Durch die Arbeit an sich selbst Sinn zu erfahren ist ein weiterer Grund, der im Mittel von den Befragten Zustimmung erhielt. Das persönliche Glück erhielt doppelt so oft die stärkste Zustimmung der Befragten wie das kurzfristige Wohlbefinden. Am dritt häufigsten[137] wurde dem Grund sich selbst kennenlernen am stärksten zugestimmt. Allen anderen im Fragebogen angegebenen Gründen, wie beispielsweise dem Vergleich mit anderen, Gruppendruck, familiäre oder gesellschaftliche Erwartungen, spirituelle Gründe, wurde im Schnitt leicht nicht bis nicht zugestimmt. Gemäss dem begrenzten Rahmen dieser Arbeit wurden hier bewusst die Gründe über alle Bereiche zusammengefasst.[138]

Die in den Interviews am häufigsten genannte Motivation für die Arbeit an sich selbst war die soziale Anerkennung und Akzeptanz - das Dazugehören[139]. 17 Personen äusserten diese Motivation in unterschiedlichen Formen: Person 3 spricht von einer "Angst, nicht mehr gebraucht zu werden in einer Gesellschaft" (Transkript, Absatz 63) und diesem "gesellschaftlichen Druck nicht allein zu sein und nicht zurückgelassen zu werden" (Transkript, Absatz 67).

[133] Eine weitere plausible Erklärung für den Unterschied in der Intensität wäre, dass im Fragebogen das an sich Arbeiten definiert wurde, während im Interview lediglich die individuellen Verständnisse der Studierenden galten.

[134] Die Antworten sind mit "Weiterentwicklung" sowie "Q2/Q3 Leben langes lernen" kodiert.

[135] In den Pretests wurde ausdrücklich gefragt, ob der Begriff im Interviewkontext als Selbstoptimierung verstanden wird. Im Fragebogen wurde zusätzlich eine Definition des Begriffs eingeblendet, worauf die Befragten mit einer Lesebestätigung reagieren mussten, bevor sie mit dem Fragebogen weiterfahren durften.

[136] Die Antworten sind kodiert mit "Vergangenheit".

[137] Sechsmal.

[138] Dies ist unter anderem möglich gewesen, weil die beliebtesten Gründe - obschon teilweise in unterschiedlicher Reihenfolge - über die Bereiche konstant waren. Zudem wurde in Bereichen, in denen im Mittel intensiver gearbeitet wird, den Gründen im Mittel stärker zugestimmt. Im Anhang sind die Gründe pro Bereich beschrieben.

[139] Hier sei zu betonen, dass in den Interviews mit der ersten Frage explizit nach der inneren Motivation gefragt wurde. Im Anschluss wurde gefragt, ob es weitere Faktoren gibt, die es erstrebenswert machen an sich zu arbeiten (was einer äusseren Motivation gleichkommt). Die entsprechenden Antworten auf beide Fragen wurden zusammen als "Q5 Anerkennung / dazugehören / Akzeptanz)" kodiert.

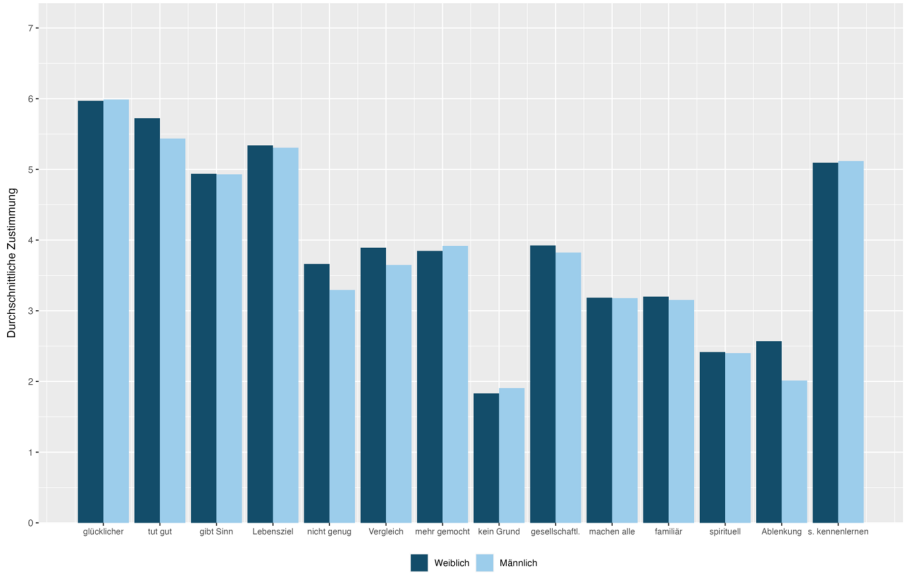


Figure 4: Durchschnittliche Zustimmung Selbstoptimierungsmotive (eigene Darstellung)

Auch Person 17 hat Respekt vor der ”Bewertung der Anderen” (Transkript, Absatz 73). Die Antworten harmonieren mit der oben beschriebenen gesellschaftlichen Erwartung, etwas aus dem eigenen Leben machen zu müssen. Aus den Antworten geht hervor, dass die Studierenden eine solche Erwartungshaltung auch aus dem engeren Umfeld erfahren. Person 10 nennt als Motivation an sich zu arbeiten ”Personen in meinem Umfeld. Ich nehme an, wenn ich an mir selber arbeite, werde ich selber oder wirke ich angenehmer auf andere. Konversationen und der Umgang wird besser” (Transkript, Absatz 67). Person 22 drückt den Wunsch nach Akzeptanz und Anerkennung folgendermassen aus: ”Ich glaube auf jeden Fall das äussere Umfeld; dass Leute stolz auf mich sind (Eltern, Verwandte oder Geschwister). Ja und auch ein bisschen das dazugehören, wenn halt alle an sich arbeiten und alle erfolgreich sind, dann will man das ja irgendwie auch (machen) – da will ich ja nicht nur faul auf dem Sofa rumliegen” (Transkript, Absatz 62). Person 24 spricht hingegen von ”Chancen” (Transkript, Absatz 57), die ohne die Arbeit an sich selbst nicht möglich wären.

Personen 21 und 23 zeigen mit ihren Aussagen auf, dass die Anerkennung und Akzeptanz auch nach innen gerichtet sein kann: ”Ich glaube ich kann mich mehr respektieren, wenn ich gewisse Dinge kann oder weiss” (Transkript, Person 21, Absatz 67) sowie ”Anspruch an mich selbst. Ich würde sagen, ich bin eher von mir selbst motiviert und froh, dass ich da keinen externen Druck brauche” (Transkript, Person 23, Absatz 81). Glückliche Beziehungen wurden von zwölf Personen[140] als Motivation genannt an sich selbst zu arbeiten. Person 2 spricht davon, ein ”nachhaltiges Umfeld aufzubauen” (Transkript, Absatz 67), während Person 6 motiviert ist Bekanntschaften zu machen, welche sie sich vor einem Jahr noch nicht getraut hätte zu machen (Transkript, Absatz 71). Andere Interviewte sind motiviert, die jetzigen Beziehungen zu pflegen: ”Und ich kann soziale Beziehungen besser pflegen, weil ich mich verhalte, wie ich das Gefühl habe, es wird erwartet”. Glückliche Beziehungen werden von Person 15[141] mit möglichst konfliktfreien Beziehungen gleichgesetzt: ”Und ich glaube, wenn man sich mit den Leuten versteht, dass mir das das Leben einfacher macht. Weil es weniger Konflikte gibt und deswegen versuche ich emotionale Intelligenz zu verbessern/ zu steigern, die Leute besser zu verstehen im Umfeld, weil das mein Leben leichter macht und weniger Konflikte gibt” (Transkript, Absatz 54).

Übereinstimmend mit den Resultaten des Fragebogens nennt mehr als ein Drittel der Befragten[142] im Interview ein Glückliches Leben als Motivation, an sich selbst zu arbeiten. Person 2 spricht davon, ”irgendwie für mein Glück selber zu sorgen versuchen” (Transkript, Absatz 67) und Person 5 ”und das andere ist sicherlich eine sehr universelle Idee; das Streben nach Glück. Ich fühl mich wirklich gut, wenn ich Dinge besser mache und wenn ich auch zum Glück von anderen und auch von mir beitragen kann, in dem ich Sachen besser mache oder Probleme löse oder neue Horizonte eröffne” (Transkript, Absatz 57). Hier sei darauf aufmerksam gemacht, dass die Befragten verschiedene Verständnisse von einem glücklichen Leben haben, auf welche später im Detail eingegangen wird.[143]

[140] Die Antworten sind kodiert mit ”Q5 Glückliche Beziehungen”.

[141] Person 23 drückt sich fast identisch aus.

[142] Insgesamt 11 Personen. Die Antworten sind kodiert mit ”Q5 Glückliches Leben”.

[143] Hier sind bereits vorab ein paar Beispiele genannt. Person 1 spricht davon, im Alter nicht zu bereuen, dass sie etwas nicht getan hat, sondern gemacht zu haben ”was ich für richtig halte, was mir wichtig ist und was mich glücklich macht” (Transkript, Absatz 56). Person 14 spricht davon möglichst viele neue Erfahrungen machen zu können (Transkript, Ab-

Weniger oft, aber dennoch regelmässig[144] wurde als Motivation für das an sich Arbeiten das Ziel, die bestmögliche Version von sich selbst sein zu wollen, genannt. Von den Studierenden wird diese Motivation folgendermassen ausgedrückt: ”Ich habe das Gefühl, wenn man mit sich selber im Reinen ist, dann kann man auch anderen viel besser helfen” (Transkript, Person 10, Absatz 68) sowie ”Und ich glaube, wir haben eine solche Verantwortung so: Du bist hier, weil du hier bist, solltest du die bestmögliche Version deiner selbst sein” (Transkript, Person 13, Absatz 59).

Viele der Personen wollen für andere oder die positive Wirkung, die sie auf andere haben können, eine bestmögliche Version von sich selbst sein. Acht Personen[145] äussern des Weiteren separat die Motivation in ihrem Umfeld etwas zu bewirken und/oder die Welt verbessern zu wollen. Die Befragten wünschen sich dies für das eigene Wohlbefinden sowie für jenes des Umfeldes und der Umwelt. Auf weitere vereinzelt genannte Motivationen für die Arbeit an sich selbst, kann aufgrund des begrenzten Rahmens dieser Arbeit nicht eingegangen werden.[146] Sofern die Studierenden bei der Motivation davon sprachen, sich selbst kennenlernen zu wollen, war es vielfach im Zusammenhang damit, dass sie mit sich im Reinen und/oder zufrieden sein wollen.[147]

Abschliessend lässt sich über die genannten Motivationen zur Selbstoptimierung folgendes sagen: Glückliche Beziehungen und ein glückliches Leben sind die dominantesten intrinsischen Motivationen für die Arbeit an sich selbst. Die Befragten sprechen noch einstimmiger von der Anerkennung und dem ‘Dazugehören’ als beliebteste extrinsische Motivation. Wie aufgezeigt, stehen die Resultate des Fragebogens nicht im Widerspruch zu jenen des Interviews. Im Fragebogen wurde zudem erfragt, wie die Studierenden die Beantwortung der Fragen zur Intensität und Motivation der Selbstoptimierung empfunden haben. Dies fiel den Befragten gesamthaft schwerer als jene Fragen zum gelingenden Leben, auf welche im nächsten Kapitel eingegangen wird. Weitere Details zur Auswertung der selbstreflexiven Frage können im Anhang C eingesehen werden.

4.4. Gelingendes Leben

Wie Abbildung 5 zeigt, stimmen die Befragten im Fragebogen im Mittel allen Aussagen zu den unterschiedlichen PERMA-Kategorien stark zu.[148] Positive Beziehungen, Sinn sowie die in dieser Arbeit hinzugefügte Kategorie ‘sich selbst kennenzulernen und weiterentwickeln können’[149] wiesen die höchsten Durchschnittswerte aus.

Es ist denkbar, dass die Fragestellung, welche die Studierenden um keine Priorisierung der Aussagen oder Kategorien gebeten hat, die hohe Zustimmung begünstigte. Spannende Muster zeichnen sich ab, wenn sich Personen zwischen verschiedenen Möglichkeiten entscheiden müssen, wie dies beispielsweise bei der Frage 6 der Fall war. In dieser wurden die Studierenden gefragt, welche der Darstellungen in Abbildung 6 für sie dem Verlauf eines gelingenden Lebens näherkommt und weshalb dies der Fall ist.

13 Personen entscheiden sich für die linke Darstellung, zehn für die rechte Darstellung, während sich fünf Personen für eine Mischung der Formen über verschiedene Zeitabschnitte oder in der Form einer Spirale oder Wellenlinie entscheiden

In den Begründungen beschreiben zwölf Personen[150] wie hier Person 13 den geraden Pfeil als unrealistisch: ”Die Vorstellung, das Leben linear zu sehen, wie es rechts dargestellt wird, denke ich, greift zu kurz. Es ist ein in sich geschlossenes komplexes, aber irgendwie sinnstiftendes Phänomen. Ich denke links. Das rechts ist mehr der Aktienmarkt” (Transkript, Absatz 73). Die meisten[151] dieser Personen entscheiden sich für die linke Darstellung. Sechs der 30 Personen vertraten die teils gegensätzliche Meinung, dass der rechte Pfeil einen Aufwärtstrend zum besseren und reiferen Selbst sei, davon wählten vier Personen auch den geradlinigen Pfeil als Darstellung eines gelingenden Lebens. Mit diesem Narrativ übereinstimmend interpretierten acht Personen die Darstellung des Kreises als stehen bleiben. Die Rundung des Kreises wird von diesen Personen als Wiederholung und nicht weiterkommen interpretiert und unter anderem mit Langeweile in Verbindung gebracht. Alle ausser eine dieser Personen entschieden sich für die rechte Darstellung. Dem gegenüber sehen 13 Personen einen Kreislauf in der Darstellung des runden Pfeils. Davon wählen acht Personen den kreisförmigen Pfeil als Repräsentation für die Darstellung eines gelingenden Lebens. Alle vier Personen, die die rechte Darstellung mit dem Wirtschaftswachstum in Verbindung bringen, wählen den Kreis als Darstellung für ein gelingendes Leben. Die restlichen Personen entscheiden sich sprunghaft zwischen den Darstellungen hin und her oder können sich für keine Darstellung entscheiden, beziehungsweise die Darstellungen für sich sinnvoll kombinieren.[152]

In der Folgefrage des Interviews wurden die Studieren-

satz 73), während sich Person 18 einfach glücklich fühlen möchte (Transkript, Absatz 57). Person 6 versucht, ein Glücksgefühl der Vergangenheit wieder zu erlangen (Transkript, Absatz 67).

[144] Es handelt sich dabei um neun Personen. Die Antworten sind kodiert mit ”Q5: Beste Version von sich selbst”.

[145] Die Antworten sind kodiert mit ”Q5: Impact / Bewirken / Einfluss / Welt verbessern”.

[146] Von acht Personen wurden Neugierde, Interesse oder die Suche nach Herausforderung als Motivation an sich zu arbeiten genannt. Mehrfach wurde auch das kurzfristige Wohlbefinden, finanzielle Ziele, Ablenkung, oder die Befriedigung durch das Erfüllen eines Ziels genannt. Weniger oft wurde eine Verantwortung, Dankbarkeit für die eigene Existenz oder das Innehahen von Privilegien als Motivation des an sich Arbeitens genannt.

[147] Entsprechend wurde kein separater Kode für sich kennenlernen kreiert, sondern die Antworten mit einem gemeinsamen Kode namens ”Q5 sich kennen / mit sich im Reinen sein” erfasst.

[148] Die Varianz innerhalb der Subfragen ist klein: Die Befragten beantworten alle Subfragen konstant hoch. Ein Boxplot Diagramm zeigt, dass die Streuung der Werte ist innerhalb der Kategorie positive Emotionen am grössten ist.

[149] Im Diagramm als ‘Entwicklung’ abgekürzt.

[150] Die Antworten sind kodiert mit ”Q6 Geradlinigkeit = unrealistisch”.

[151] Es handelt sich um eine knappe Mehrheit von sieben Personen.

[152] Manche sehen in beiden Darstellungen gelingende Leben und beschreiben, dass dies je nach Individuum unterschiedlich ist.

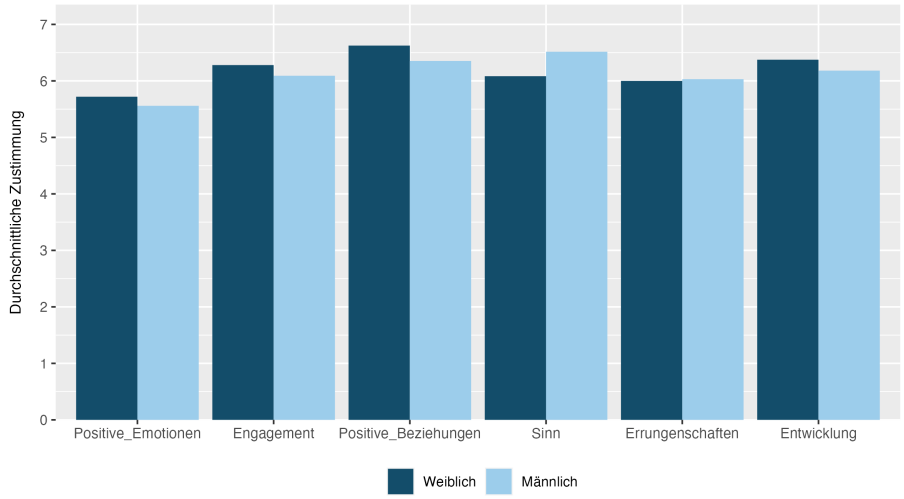


Figure 5: Vorstellungen eines gelingenden Lebens anhand des PERMA-Modells und die Bedeutung von Entwicklung darin (eigene Darstellung)

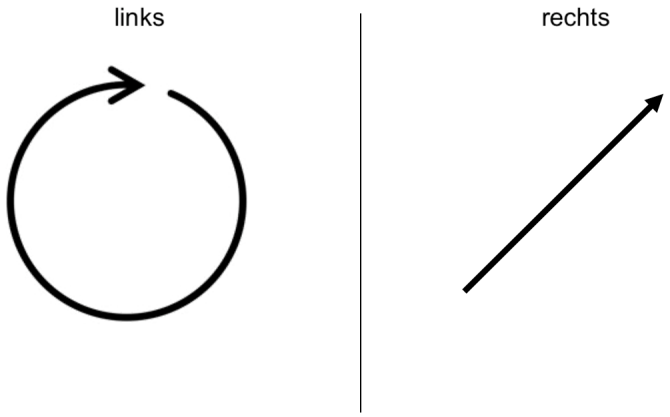


Figure 6: Auswahl Formen eines gelingendes Leben (eigene Darstellung)

den gefragt, was für sie, abgesehen von den Darstellungen, ein erfülltes Leben ausmacht.

Positive Beziehungen wurde von 20 Personen als massgebend für ein erfülltes Leben genannt und stellt somit die häufigste Antwort dar. Die Befragten verstehen darunter "geliebt werden und lieben" (Transkript, Person 2, Absatz 87) oder eine eigene Familie zu gründen. Person 10 spricht von einem wichtigen "Tiefgang" (Transkript, Absatz 81), der durch Beziehungen möglich wird, und das Zitat von Person 3 veranschaulicht, dass viele der Befragten Verbundenheit und Tiefe in ihren direkten menschlichen Interaktionen verspüren möchten: "Dann gute Beziehungen mit Freunden, Familie, Partner usw. und mit guten Beziehungen meine ich wirklich so Qualität über Quantität. Lieber weniger aber dafür tiefere Beziehungen" (Transkript, Absatz 83).

Auf die Frage nach dem gelingenden Leben nennen zehn Personen[153] die Erfüllung von Grundbedürfnissen wie Essen, Trinken sowie finanzielle Sicherheit. Für ebenfalls zehn Personen[154] gehört das Empfinden positiver Gefühle zu einem gelingenden Leben. Die Interviewten nennen Spass und Freude am Leben haben, die Welt bereisen sowie neue spannende Erfahrungen machen in diesem Bereich.

Zehn Personen[155] nennen sich selbst kennen oder mit sich im Reinen sein als essenziellen Teil eines gelingenden Lebens. Person 1 fasst dies folgendermassen in Worte: "Dass man nicht ein fremd bestimmtes Leben führt, sondern bei sich selbst ist, sich selbst kennenlernt und dann auch bei sich selbst ist und eigentlich, dass man seine tiefsten Überzeugungen stärken, Wünsche und Antriebe verwirklichen kann [...]

[153] Die Antworten sind kodiert mit "Q7 Grundbedürfnisse".
[154] Die Antworten sind kodiert mit "Q7 Positive Emotionen".
[155] Die Antworten sind kodiert mit "Q7 Sich kennen / mit sich im Reinen sein".

Ja, losgelöst von dem, was andere denken und losgelöst von dem, was sein soll - mehr was für einen selbst sein möchte" (Transkript, Absatz 75). Auch Person 10 beschreibt das sich selbst Kennen ähnlich heldenhaft: "Also ich glaube ein gelingendes Leben ist im Einklang zu leben mit sich selber und dem, was um einen herum passiert, also seine Bedürfnisse zu verstehen und diese dann entsprechend zu befriedigen, wie es einen dann glücklich macht. Ich glaube, das ist so zusammengefasst, das, was ein gelingendes Leben ist. Wie gesagt, das ist sehr, sehr stark, abhängig davon, dass man sich selber gut kennt" (Transkript, Absatz 81). Andere Personen beschreiben das mit sich im Reinen sein als nichts bereuen zu wollen (Transkript, Person 11, Absatz 82) oder nicht unnötiges Leid verursachen (Transkript, Person 2, Absatz 87).

Engagement wird von neun Personen[156] als Teil eines gelingenden Lebens gesehen. Die Studierenden bekunden Freude daran herausgefordert zu werden, dem eigenen Interesse nachzugehen und die eigene Neugierde befriedigen zu können. Das Engagement wird insbesondere auf die Arbeit oder das Studium bezogen.

Ähnliche Tendenzen sind bei der Nennung von Sinn zu verzeichnen. Wie im nächsten Kapitel verdeutlicht, wird Sinn von den Studierenden oft genannt, selten aber ausführlich beschrieben. Vielfach wird Sinn in Kombination mit einer beruflichen Tätigkeit genannt, die sowohl herausfordernd als auch sinnvoll sein soll[157]. Sinn wurde wie angedeutet vielfach mit dem in Verbindung gebracht, mit dem, "was man macht" (Transkript, Person 4, Absatz 73). Für die einen stellt das einen Beitrag zur Nachhaltigkeit dar, für die anderen einen für die 'community' (Transkript, Person 6, Absatz 86).

Unter dem Kode "Q7 Andere" wurden weitere weniger häufig genannte Bestandteile eines gelingenden Lebens wie Achtsamkeit, Bescheidenheit, Autonomie gesammelt, auf welche in dieser Arbeit aufgrund des begrenzten Rahmens nicht weiter eingegangen wird.

Auf die Folgefrage, ob die Studierenden etwas tun, um das zu erreichen, was sie zuvor als gelingendes Leben beschrieben haben, antworteten die allermeisten Studierenden mit Tätigkeiten, die denen der vierten Frage ähneln: 20 Personen[158] nennen das aktive sich Zeit nehmen für Beziehungen, neue Kontakte knüpfen und pflegen als Tätigkeit. Zwölf Personen[159] nennen sich Wissen aneignen und lernen als eine Tätigkeit. Mehr als die Hälfte der Personen bezieht dies auf das aktuelle Studium. Es geht aber auch um das Wahren einer Balance der Interessen.[160]

Sich selbst kennenzulernen, wird von acht Personen als Tätigkeit genannt. Dazu zählen die Befragten, innere Stärke aufzubauen sowie möglichst viele verschiedene Perspektiven kennenzulernen um zu sehen, welche einem am meisten liegt. Neu findet der Kode "Selbstsorge, Zeit für sich" bei acht Personen[161] Anwendung. Die Befragten beschreiben unter anderem, wie sie ihr eigenes Wohlbefinden wieder auf den Radar bringen, bewusst im Moment sein oder starke Leistungserwartungen ablegen möchten. Die Zeit für sich ist vielfach auch an Zeit für Selbstreflexion geknüpft. Selbstreflexion als Methode wurde von neun Personen genannt.

Grundsätzlich kann gesagt werden, dass in ähnlichen Bereichen, entsprechend den vorherrschenden Vorstellungen eines gelingenden Lebens, (an sich) gearbeitet wird. Wenige Personen[162] äusserten das Gefühl, dass das Erreichen eines gelingenden Lebens, wie sie es sich vorstellen, nicht kontrollierbar ist. Andere[163] überkommt das Gefühl, dass sie für das gelingende Leben noch mehr arbeiten müssten.

Die Rückmeldungen auf die letzte Interviewfrage, die die Studierenden nach ihren persönlichen Zusammenhängen zwischen den Antworten auf die Fragen des an sich Arbeitens und des gelingenden Lebens fragt, bestätigt den Link zwischen Arbeit an sich selbst und dem gelingenden Leben: 20 Personen bestätigen in unterschiedlicher Intensität, dass sie für ein gelingendes Leben an sich arbeiten. Acht Personen bemerken die Überschneidungen in ihren Antworten, stellen aber sonst keinen Zusammenhang bewusst her. Während eine Person die Frage nicht ausreichend[164] beantwortet, sieht eine Person keinen Zusammenhang. Es überwiegt die Erzählung, dass durch die Arbeit an sich selbst bewusst auf ein gelingendes Leben hingearbeitet wird. Entsprechend wurden die Antworten des Fragebogens darauf überprüft, inwiefern einzelne Faktoren eines gelingenden Lebens als Selbstoptimierungsgründe angeführt werden und ob dies beispielsweise mit der Wichtigkeit dieser Faktoren einhergeht.[165]

Die Forschungsfrage dieser Arbeit lautete, inwiefern Selbstoptimierungsnarrative und -praktiken mit Heldenreisen konkurrieren. Um die Frage beantworten zu können, wird im nächsten Kapitel auf die Resultate der Heldenreisen eingegangen.

[156] Die Antworten sind kodiert mit "Q7 Engagement".

[157] Für Tätigkeiten, die von den Studierenden als sinnvoll beschrieben wurden, wurde ein neuer Kode "Q7 erfüllende Tätigkeit" entwickelt. Dieser und jener für den Sinn und Zweck im PERMA-Modell wurden je sieben Mal – zum Teil auch überschneidend - verwendet.

[158] Die Antworten sind kodiert mit "Q8: Methode Social skills / Austausch / Netzwerk".

[159] Die Antworten sind kodiert mit "Q8 Methode Wissen / skills aneignen, lesen".

[160] Es wird besonders im sozialen Bereich (12 Personen), im Beruflichen (11 Personen), Akademischen (7 Personen) sowie der Gesundheit für ein gelingendes Leben gearbeitet. Die Antworten sind kodiert mit "Q8 Bereich Soziales", "Q8 Bereich Beruflich", "Q8 Bereich Akademisch", "Q8 Bereich Gesundheit".

[161] Die Antworten sind kodiert mit "Q8 Methode Selfcare / Zeit für sich". Der Kode, der in der Frage 4 "eigenes Glück / Persönlichkeit" oft Anwendung fand, wurde in dieser Frage möglicherweise auch wegen dem neuen Selbstsorgekode nur viermal angewendet.

[162] Die Antworten sind kodiert mit "Q8: Keine Kontrolle".

[163] Die Antworten sind kodiert mit "Q8: Gefühl, mehr machen zu müssen".

[164] Die Person geht bei ihrer Antwort nicht ausreichend auf die gestellte Frage ein, was der Interviewerin erst beim Transkribieren, nicht aber in Echtzeit, auffiel.

[165] Dazu wurden die Zustimmung zu den PERMA-Kategorien und dem ergänzten 'sich selbst kennenlernen' mit den dazu passenden Selbstoptimierungsgründen untersucht. Die Beschreibung dessen kann im Anhang C eingesehen werden.

4.5. Heldenreisen

Die Auswertung der Interviews ergab, dass sich eine Minorität der Befragten[166] auf einer inneren Heldenreise befindet. Während eine Person nur eine innere Heldenreise erlebt, konnte bei fünf Personen sowohl eine innere als auch eine äussere Heldenreise festgestellt werden. Die Hälfte der Befragten befindet sich ausschliesslich auf äusseren Heldenreisen. Bei den restlichen neun Interviewten konnte keine Heldenreisen-ähnliche Struktur aus den Erzählungen festgestellt werden. Die Analyse der Personen, die sich auf Heldenreisen befinden, ergab zudem, dass sich Männer im Vergleich zu Frauen tendenziell weniger auf innere Heldenreisen begeben. Personen, die den Kurs Ökonomie des Glücks besucht haben, tendieren dazu, sich auf Heldenreisen zu begeben oder zu befinden. Beim Glauben ist kein Muster erkennbar.[167]

In der Abbildung 7 wurden die Summen der angegebenen Intensität des an sich Arbeitens[168] aller Personen in den jeweiligen Heldenreisenkategorien abgebildet[169]. Die Darstellung veranschaulicht, dass ein hohes Mass an Selbstoptimierung keine Art der Heldenreise per se ausschliesst[170]. Es ist interessant zu sehen, dass die höchste Summe der Selbstoptimierung von der Person erreicht wurde, die eine innere Heldenreise macht. Aufgrund der geringen Anzahl interviewter Personen und dem primär qualitativen Ansatz zur Auswertung der Heldenreisen lassen sich keine aussagekräftigen quantitativen Resultate ausweisen. Infolgedessen liegt der Fokus auf den erkannten Mustern der qualitativen Analyse.

Die nächsten Abschnitte sind den Inhalten der Interviews gewidmet, auf denen die Einteilung in die vier Kategorien innere Heldenreise, äussere Heldenreise, innere und äussere Heldenreise sowie keine Heldenreise basiert[171].

Zuerst werden Trends beschrieben, die auf eine äussere Heldenreise hindeuten.[172] Wie bereits in den Kapiteln zum gelingenden Leben und den Selbstoptimierungsmotivationen behandelt, sprechen insgesamt 14 Personen davon, etwas bewirken zu wollen oder sich aktivistisch für etwas einzusetzen. Die Befragten sprechen davon "einen Unterschied zu machen" oder "etwas Gutes" zu tun. Der Akt des sich Auflehnens ist an sich nicht heroisch, sondern kann quasi heroische Züge annehmen.[173]

19 Personen berichten davon, dass sie daran arbeiten, ihre Emotionen oder Triebe zu kontrollieren oder erzählen davon sich selbst zu betäuben. Dies wurde im Rahmen dieser Arbeit als nicht heroisch angesehen, weil die Kontrolle von Emotionen zu einem gewissen Grade impliziert, dass die Personen ihre Emotionen nicht spüren und aushalten können, sowie diese möglicherweise nicht im Sinne ihres wahren Selbst ausleben können. Das Betäuben kann beispielsweise in Form von Überarbeitung auftreten und dazu führen, dass Emotionen und Bedürfnisse gar nicht erst spürbar werden. Personen 3 und 23 beschreiben die Kontrolle von Emotionen folgendermassen: "Also ich habe jetzt so vor 1-2 Monaten angefangen, (fast) jeden Tag zu meditieren (das habe ich so mit einer App gemacht) und auch so eine mindfulness Meditation jeden Tag, die so 10 Minuten geht. Das ist so bisschen eine Art mich zu 'grounden'. So versuchen meine emotionale Seite zu zähmen zu wissen, und dass die mir nicht alles ruiniert" (Transkript, Person 3, Absatz 59) sowie

> "Ich denke persönlich ist die Motivation eher ein harmonisiertes Leben/Umfeld haben zu können, wo man sich nicht ständig mit Konflikten oder negativen Gefühlen herumschlägt. Man möchte ja auch eine Verbindung zu anderen Menschen haben. Ich denke, egal wie individuell ein Mensch ist, man braucht trotzdem ein Minimum an sozialen Kontakten/Umfeld, um glücklich zu sein und da ist der Trieb in einem einfach tief verankert, ein Umfeld zu haben und akzeptiert zu werden und da kann man sich ja nicht verhalten, wie man will" (Transkript, Person 23, Absatz 75).

Folgende Aussagen enthalten Hinweise, dass sich zu betäuben unter den Befragten ein gängiges Narrativ ist: "Ich glaube ich bin jemand, der immer beschäftigt sein muss. Ich kann irgendwie nicht immer das Gleiche machen. Ich finde es mega spannend, wenn jeder Monat und jedes Jahr anders aussieht. Ich finde den Einblick in verschiedene Dinge des Lebens sehr spannend und man sollte sich nicht auf irgendetwas begrenzen, sondern sich wirklich weiterentwickeln" (Transkript, Person 4, Absatz 56). "Ich merke, wenn ich so gar nichts zu tun habe im Leben, wird mir langweilig" (Transkript, Person 7, Absatz 60). Die Rückmeldungen spiegeln die Befürchtung gewisser Befragter, 'sich im Kreis zu drehen'[174]. Das Betäuben kommt der Zielerreichung ohne Ziel aus der Theorie[175] gleich. Person 9 beschreibt die Tatsache, dass sie in den Semesterferien aufgrund von

[166] Es handelt sich um sechs Personen, wobei sich fünf Personen zeitgleich auch auf einer äusseren Heldenreise befinden.

[167] Aufgrund des niedrigen N kann nicht von signifikanten Trends gesprochen werden. Die genannten Tendenzen müssten mit einer grösseren Stichprobenmenge überprüft werden. Im Anhang C können ausgewählte Übersichtsdarstellungen betrachtet werden.

[168] Gemessen anhand der Intensität der Ausführung der Tätigkeiten über alle Bereiche.

[169] Eine vergleichbare Darstellung wurde mit der Summe der Zustimmung der Gründe für die Selbstoptimierungstätigkeiten entworfen und kann im Anhang C eingesehen werden.

[170] Sowie umgekehrt.

[171] Die genannten vier Kategorien sind im Kodierleitfaden sogenannte Fazitkodes. Im Kodierleitfaden wurden während des Auswertens sogenannte Indikatorenkodes erstellt und angewendet, die eine gewisse Art der Heldenreise indizieren. Die Beschreibung der Trends in den Indikatorenkodes zeigt ein holistischeres Bild der Heldenreisen(-anteilen) der Befragten. Mit den nachfolgenden Links können alle angewendeten Fazit- und Indikatorenkodes eingesehen werden.

[172] Es sei hier darauf hingewiesen, dass die Indikatorenkodes nicht in jedem Fall voneinander abgrenzbar sind beziehungsweise je nach dem sogar miteinander in Verbindung stehen.

[173] Die Theorie dazu kann im Kapitel 2.3.3 eingesehen werden.

[174] Referenz zu den Rückmeldungen auf die Interviewfrage 6; welche der beiden Darstellungen dem gelingenden Leben näherkommt und wieso.

[175] Die Theorie dazu befindet sich im Kapitel 2.3.3.

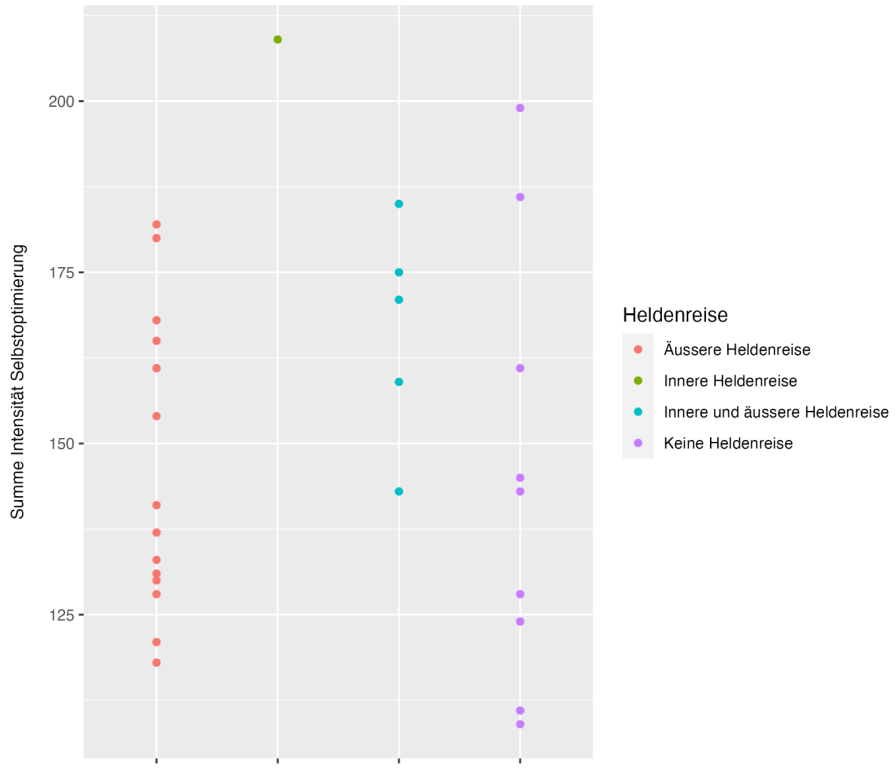


Figure 7: Totale Selbstoptimierungsintensität und Art der Heldenreise (eigene Darstellung)

Langeweile bereits wieder Bücher für die Universität liest als ”toxischen Selbsthass” (Transkript, Absatz 68).

18 Personen sprachen in den Interviews zudem von Lebensprojekten, die sie stark mit der eigenen Identität und dem eigenen Glück verknüpfen. Es konnten drei verschiedene Arten von Lebensprojekten ausgemacht werden, die vermeintlich zu Glück führen.[176] Erstens scheinen die Befragten gute Beziehungen mit ihren Mitmenschen sowie die Gründung einer Familie als Lebensprojekte zu sehen. Folgende zwei Zitate illustrieren dies: ”Was für mich auch sehr wichtig wäre - das habe ich glaube ich noch nicht so gesagt - ist, dass ich schon eine Familie haben und eine sehr gute Beziehung zu meinem Mann und meinen Kindern haben möchte. Das ist etwas, was mir wahrscheinlich auch sehr viel Zufriedenheit geben kann, so über eine längere Zeit einen ‘happy’ machen kann” (Transkript, Person 20, Absatz 104). Sowie ”ein anderes Ziel (da gibt es natürlich Etappen) ist Familie: eine eigene Familie gründen könnte ein Ziel sein und ist auch ein Ziel von mir” (Transkript, Person 8, Absatz 91). Person 26 spricht sogar vom Ziel, sich nicht von der Partnerin trennen zu wollen (Transkript, Absatz 64).

Zweitens zeichnet sich ab, dass einige Studierende eine erfüllende Tätigkeit als Lebensprojekt ansehen. Das kann aktuell das Studium sein oder auch der ganze absehbare Karriereweg. Person 8 möchte beispielsweise, verbunden mit dem Berufsleben, viel reisen (Transkript, Absatz 92), während Person 15 viel Geld verdienen möchte, um sich viel leisten und viel erleben zu können:

> ”Deswegen auch der starke Fokus auf die Karriere, weil ich weiss okay, mit einem normalen Job wird man nicht unbedingt reich werden – es würde zwar zum Leben reichen, aber ich mag zum Beispiel auch sehr gerne Autos; also so Ferraris finde ich sehr schön, so von der Form. Ich bin da sehr fasziniert von und das kriegt man auch nicht mit ‘nem normalen Job und deswegen weiss ich, dass ich für die Karriere da bisschen mehr geben muss als vielleicht andere” (Transkript, Absatz 61).

Drei Personen[177] sind zudem fasziniert von der Verantwortung und Rolle als Unternehmer*in oder Manager*in.

Die beste Version von sich selbst sein zu wollen und folglich konstant an sich zu arbeiten, konnte als dritte Lebensprojektkategorie identifiziert werden. Person 5 beschreibt dies folgendermassen:

> ”Ich habe einfach den inneren Willen, ein Leben lang (über) an mir zu arbeiten, in dem Sinne, dass niemand perfekt ist. [...] Es geht nicht

[176] Die Theorie dazu kann im Kapitel 2.1 eingesehen werden.

[177] (Tanskript, Personen 12, Absatz 81), (Transkript, Person 15, Absatz 79), (Transkript, Person 30, Absatz 95).

> darum, perfekt zu sein oder irgendwelche Ideale, die vielleicht auch krank sind oder krank machen können, zu verfolgen, sondern einfach, dass man Fehler sieht, dass man Dinge sieht, die man besser machen hätte können, auch gegenüber anderen, und dass man daraus lernt. Man lernt ein Leben lang, und meine Einstellung ist es, dass das wichtig ist, auch für mich persönlich" (Transkript, Absatz 47).

Das Suchen des Glücks in Lebensprojekten kommt dem Kreisen um die unbeschreibbare und unerreichbare Mitte nah. Folgende Zitate machen deutlich, dass dieses Kreisen bei knapp der Hälfte der Befragten, vielfach im Zusammenhang mit äusseren Heldenreisen, ein Thema ist. Beispielhaft sieht Person 3 Sinn im Leben in der Arbeit: "Ich hoff mal später irgendwo zu arbeiten, wo mich die Arbeit erfüllt und einen gewissen Sinn erweckt" (Transkript, Absatz 87). Person 15 strebt danach, Glück durch Wohlstand zu erreichen: "Ja schon, also ich meine, wenn man ein gutes Leben hat oder in einem schönen Auto sitzt - also ich bin jemand - ich würde auch in der Stadt ein bisschen ‚rumposen' gehen, weil ich's auch lustig finde, mit 'nem Freund zusammen oder so. Das ist auch ein Teil. Man möchte den Leuten ja auch zeigen ‚ich bin hier und ich hab's geschafft'" (Transkript, Absatz 67). Sinn wird auch von Person 14 nicht im Sinne der inneren Heldenreise, als Teil eines Grösseren, sondern wie folgt beschrieben: "Eben wertstiftend zu sein für andere. Die Möglichkeiten haben. Ich glaube bei mir ist auch recht stark präsent so 'Ich muss jetzt wachsen, damit ich ein Leben für mich aufbauen kann'. Eben auf der einen Seite finanziell, aber auch auf anderen Seiten; Beziehungs-technisch" (Transkript, Absatz 69).

Person 7 ist sich hingegen bewusst, dass sie im Leben noch auf der Suche ist: "Da ich da selbst noch im Leben am Suchen bin und das noch nicht so ganz gefestigt ist, ist es schwierig zu sagen, aber jetzt auf die längere Zukunft, wäre mir das schon auch glaube ich wichtig jemand zu haben, mit dem ich das Leben teilen kann, eine eigene Familie zu gründen" (Transkript, Absatz 81). Wie bereits in vorherigen Kapiteln beschrieben sind einige Befragte dabei, sich aktiv kennenzulernen und streben Erfahrungen an, die ihnen ermöglichen, sich kennenzulernen. Während oberflächliche Narrative um das sich kennenlernen existieren, sieht Person 1 Folgendes darin:

> "Zurückfinden zum Urvertrauen, zu unseren Urstärken und zu dem was (man so zu sagen) wirklich ist. Was einen wirklich glücklich macht, bevor man sozusagen geprägt wurde von allen Einflüssen, die vielleicht nur zu Reaktionen geführt haben, statt Aktionen und ich glaub so, dass für mich ist es stark ein Zurückfinden zu 'wer man wirklich ist' und dann das leben zu können. Also in einem ersten Schritt herausfinden zu können, wer man ist, und das dann leben können, ja vielleicht wie eine Heldenreise. Auch so in dem Beispiel, dass man mal runter geht und abtaucht in diese Tiefen des Selbst, es dann effektiv ausleben kann, so dass zum Schluss halt auch so ein gängiger Begriff in der Sprache ist, schlüssig ist. Also, dass das Leben in sich schlüssig ist" (Transkript, Absatz 69).

Narrative des sich verstehen und oder verändern Wollens sind entsprechend in Kontexten der inneren sowie der äusseren Heldenreise aufgetreten: Person 6 erlernt im Moment zu sein, während Person 2 zu sich selbst gut zu sein lernt. Person 3 verändert sich in Richtung mehr Resilienz. Einige Befragte arbeiten daran alte Glaubenssätze, die bekannte Welt hinter sich zu lassen. Person 17 spricht sogar davon "aus alten Mustern herauszukommen" (Transkript, Absatz 80). Bei wenigen Interviewten kommt sogar der Kode Kampf gegen die eigenen Dämonen zu Anwendung, wie bei Person 24, die den Umgang mit den eigenen Angststörungen und Depressionen erlernt.[178] Nur eine Person äusserte sich so, dass sie als Person von zwei Welten identifiziert werden konnte.

Die bisher beschriebenen Resultate haben gezeigt, dass sich in den Interviewantworten einiger Personen keine ausreichend Heldenreisen-ähnlichen Struktur erkennen lassen. Bei den Befragten, bei welchen eine transformative Struktur erkennbar war, dominierte die äussere Heldenreise über die innere. Intensive Arbeit an sich selbst verunmöglicht keine innere Heldenreise. Es scheint von grösserer Bedeutung zu sein aus welcher Motivation und daraus abgeleitet wie an sich gearbeitet wird. In anderen Worten; ob ein tieferes ‚wozu' existiert, das die Arbeit an sich selbst stützt.

Um auf die Forschungsfrage dieser Arbeit zurückzukehren, ist es entsprechend möglich, dass Selbstoptimierungsnarrative und -praktiken neben inneren Heldenreisen existieren und/oder ausgeführt werden. Falls in Lebensprojekten, wie beispielsweise der kontinuierlichen Arbeit an sich selbst, Glück erreichbar scheint, können tiefergreifende innere Heldenreisen durch äussere verdrängt werden. Die vorläufige Antwort auf die Forschungsfrage soll im nächsten Kapitel ausführlich, unter Einbezug weiterer Forschung und im Rahmen des säkularen Gesellschaftsbildes diskutiert werden. In einem ersten Schritt werden das von den befragten Studierenden wahrgenommene Gesellschaftsbild sowie die darin enthaltenen Sinn- und Selbstoptimierungsnarrative diskutiert. Im Anschluss werden die Selbstoptimierungspraktiken und -gründe der Studierenden mit anderen Studien verglichen. Schliesslich werden die identifizierten Zusammenhänge der Vorstellungen eines gelingenden Lebens und der Selbstoptimierung der Studierenden mit bestehender Literatur verglichen und kritisch diskutiert. Das Kapitel schliesst

[178] Zur Anwendung von Kodes der inneren Heldenreise ist ein Verunsicherungsmoment der eigenen Identität essenziell. Zur Vereinfachung der Auswertung wurde auch ein Kode "Vergangenheit" erstellt, der diese Abgrenzung verdeutlicht. Kolmar (2021, S. 208 ff) beschreibt, dass gewisse Transformationsprozesse in unserer Gesellschaft auf psychische Krankheiten reduziert werden. Damit sollen psychische Krankheiten keineswegs romantisiert werden, sondern lediglich auf das vielfach übersehene Transformationspotential dieser aufmerksam gemacht werden.

mit Vorschlägen zum künftigen Umgang mit den erhaltenen Resultaten.

5. Diskussion

5.1. Die Arbeit an sich selbst in einer säkularen postheroischen Gesellschaft

Die befragten Studierenden bezeichnen die Gesellschaft nicht durchgehend als säkular oder postheroisch. Heroisch scheint den Befragten allgemein kein Begriff für eine Gesellschaft zu sein - das Wort wurde nie verwendet. Modern oder säkular wird von wenigen Befragten als Bezeichnung für die Schweizer Gesellschaft verwendet. Das Bild einer säkularen Gesellschaft gemäss C. Taylor (2007) ergibt sich aus den fehlenden Bezügen auf einen gemeinsamen Glauben oder ein vergleichbar Sinn und Moral gebendes Konstrukt. Das Gesellschaftsbild, das sich aus den Interviews ergibt, stimmt überein mit jenem der Forschenden, die ein postheroisches, säkulares Zeitalter proklamieren (Campbell, 1953; Kolmar, 2021, S. 61; C. Taylor, 2007).

Die Narrative der eigenen Leistung und Arbeit an sich selbst ziehen sich durch die Interviews und den Fragebogen. Die Schweiz wird als leistungs- und zielorientiert, manchmal gar kompetitiv beschrieben. Während auf die Frage nach den gesellschaftlichen Leistungserwartungen die Antworten vorerst zurückhaltend waren, spricht sich eine deutliche Mehrheit für die Wichtigkeit und Erwartung der Arbeit an sich selbst aus. Ob die hohe individuelle Leistungserwartung im Rahmen dieser Bachelorarbeit positiv oder negativ bewertet wird, hängt insbesondere davon ab, ob die Arbeit an sich selbst aus einer Wie-Frage entspringt oder Konsequenz einer sinnstiftenden Wozu Frage ist. Die Arbeit an sich selbst ist, in anderen Worten, nach dem dahinterstehenden Sinn zu bewerten, worauf im Verlauf der Diskussion weiter eingegangen wird.

Die Frage nach dem eigenen Platz in der Gesellschaft wird wiederholt über die eigene Leistung, dem eigenen Beitrag zur Gesellschaft, beantwortet. Das Gefühl von Zugehörigkeit wird mehrfach darüber definiert, ein funktionierender Teil der Gesellschaft zu sein. Duttweiler beschreibt eine vergleichbare Verbindung ”an sich selbst zu arbeiten, seine Schwächen auszugleichen und seine Stärken auszubauen, ist heute eine Schlüsselkompetenz der gesellschaftlichen Integration geworden” (Duttweiler, 2013, S. 95). Auch Bröckling beschreibt in seinem Buch (2013), dass unternehmerisch handelnde Individuen durch Selbstoptimierung und Erfolgsstreben in einer dynamischen Welt zu bestehen versuchen.

Einige Befragte drücken die implizite gesellschaftliche Leistungserwartung als Angst ‘abgehängt zu werden’ aus. Besonders im beruflichen Kontext konnte eine solche Angst bereits in anderen Studien (Aichinger & Bauer, 2014; Hannich & Berger, 2011; Lengfeld, 2019) ausgewiesen werden[179]. Verglichen mit bestehender Literatur, zum Beispiel (Albert et al., 2016, S. 243; Aichinger and Bauer, 2014), welche die Mittvierziger sowie Jugendliche aus unteren sozialen Schichten als gefährdete Personen betrachtet, sind die Befragten dieser Bachelorarbeit erstaunlich jung und aus einer hohen sozialen Schicht.

Die Bedeutung, die der eigenen Leistung und Arbeit an sich selbst beigemessen wird, lässt die Vermutung zu, dass die befragten Studierenden dazu tendieren, fehlende grössere Sinnnarrative mit dem Imperativ ‘an sich selbst zu arbeiten zu müssen’ zu füllen versuchen. Obwohl diese These in weiteren Studien geprüft werden müsste, ist ein Vergleich mit anderen analogen Thesen an dieser Stelle wertvoll. In der Studie von Mühlhausen (2016a, S. 8) wird beispielsweise davon gesprochen, dass die Selbstverantwortung das Schicksal besiegt: ”Nachdem alle Ideologien ausgedient haben, bleibt dem freien Menschen bloss diese eine grosse Metaidee: Mach das Beste aus dem eigenen Leben”. Das Beste aus dem eigenen Leben und sich selbst machen zu wollen, wurde von knapp einem Drittel der Befragten als Motivation für die Arbeit an sich selbst genannt.

Rebentisch (2012, S. 172) argumentiert, dass Selbstverwirklichung der neue Religionsersatz ist: ”Man muss nur wollen. Eigenverantwortung, Initiative, Flexibilität, Beweglichkeit werden so zu Forderungen, die zwar das alte Disziplinarmodell der Gesellschaft verdrängt haben, ohne dabei jedoch die Einzelnen von der Selbstdisziplin zu befreien. An die Stelle einer Normierung des Subjekts nach gesellschaftlich vorgeschriebenen Rollen ist die Erwartung der kreativen Selbstverwirklichung getreten”. Girkinger (2012, S. 277, S. 20) argumentiert, dass Persönlichkeitsentwicklung im Kontext der Rivalitätslogik eine Anpassungsleistung darstellt, um gesellschaftlich mitzuhalten. Persönlichkeitsentwicklung und lebenslanges Lernen sind ”die einzig verbliebenen Krisenbewältigungsstrategien”.

Narrative, wie das ‘Leben lange Lernen’ sowie ‘sich weiterzuentwickeln’, wie sie in den Interviews für diese Bachelorarbeit genannt wurden, haben gesellschaftlich an Aufschwung erhalten und laut Girkinger (2012, S. 216) mehr Raum im öffentlichen Diskurs eingenommen.

Wie genau die Studierenden von der Arbeit an sich selbst erzählen, wird im nächsten Kapitel diskutiert. Es lässt sich festhalten, dass ein Grossteil der befragten Studierenden in der säkular und postheroisch wahrgenommenen Gesellschaft, entsprechend gesellschaftlicher Leistungsnarrative, ein Gefühl verinnerlicht hat, dass im eigenen Leben immer noch ”mehr geht” (Girkinger, 2019, S. 4) und dies möglicherweise für gewisse Befragte fehlende grössere Sinnnarrative ersetzt und so ein Zugehörigkeitsgefühl entstehen kann.

5.2. Erzählungen von Selbstoptimierung

Wie mehrfach in der Arbeit beschrieben, besitzen die befragten Studierenden starke Narrative rund um die Arbeit an sich selbst: Mehr als 90 % der Studierenden berichten von der eigenen Erwartungshaltung, an sich zu arbeiten. Die Befragten scheinen sich dem ”Zeitgeist der Selbstoptimierung”

[179] Obschon laut Lengfeld (2019) diese Angst in Deutschland im Jahr 2016 einen Tiefstand erreichte. Die Angst erinnert auch an das von Bröckling (2013) beschriebene ‘erschöpfte Selbst’.

(Girkinger, 2019, S. 4) angepasst zu haben. In Deutschland bekennen sich laut einer Studie zur Selbstoptimierung im Gesundheitsbereich (Mühlhausen, 2016a) 59% der Befragten dazu, sich mehr oder weniger selbst zu optimieren[180]. Die Studie zeigt zudem, dass sich die jungen Personen am meisten mit dem Begriff Selbstoptimierung anfreunden können (S. 30), was zu den Zahlen dieser Arbeit passt, in welcher der Begriff nicht verwendet wurde. Die Personen, die das Forschungsdesign dieser Arbeit testeten, gaben an, dass sie an sich arbeiten im Kontext der Befragungen durchaus als Selbstoptimierung verstanden haben.

Unter den befragten Studierenden herrscht grosse Einigkeit darüber, dass die Arbeit an sich selbst erstrebenswert ist. Hinsichtlich dessen, was die Selbstoptimierung umfasst, in welchen Bereichen und wie intensiv diese betrieben wird, liegt Heterogenität vor. Hinweise darauf sind unter anderem in der Diversität der Selbstoptimierungsverständnisse zu finden. Für die starke Erwartungshaltung, an sich selbst zu arbeiten, gaben die Individuen im Fragebogen unerwartet tiefe Intensitäten an, mit welcher sie Selbstoptimierungspraktiken ausführen. Einerseits kann dies daran liegen, dass Intensität kein gewohnter Massstab war, Selbstoptimierungspraktiken zu bewerten. Zudem kann es, wie im Resultate-Teil angedeutet sein, dass die angebotenen Tätigkeiten nicht auf die Personen gepasst haben. Es ist zudem plausibel, dass die im Fragebogen enthaltene Definition der Arbeit an sich selbst zur Zurückhaltung der Studierenden beigetragen hat. Zusammen mit den verschiedenen Selbstoptimierungsverständnissen kann darauf geschlossen werden, dass Selbstoptimierung für die Interviewten eine höchst individuelle Angelegenheit ist und deren Inhalt, wie in der Theorie beschrieben, beliebig füllbar ist (Röcke, 2021, S. 178).

Die erkennbaren Trends hinsichtlich der Bereiche und Methoden der Selbstoptimierung werden in der Folge mit anderen Studien und im Kontext der Theorie diskutiert.

Die Studie von Mühlhausen (2016a) weist Selbstoptimierungspraktiken und -narrative im Bereich Gesundheit in Deutschland aus. Gesundheit stellt auch in dieser Bachelorarbeit einen mehrfach genannten Bereich der Selbstoptimierung dar und wird wie in der Studie mit Wohlbefinden und Glück verbunden. Sport sowie gesunde Ernährung sind Methoden, die in beiden Forschungen genannt werden. Das daraus resultierende Erscheinungsbild ist kaum die Motivation. In den Interviews wird das Aussehen gar nicht als Bereich genannt. In der Studie von Mühlhauser wird argumentiert, dass sportlich das neue attraktiv darstellt (S. 17).

Mehr als an der Gesundheit, sozialen Beziehungen und der eigenen Persönlichkeit arbeiten die Befragten im beruflichen und akademischen Bereich an sich. Dass beide Bereiche jeweils von fast 2/3 der Befragten genannt werden, kann wahrscheinlich auch damit erklärt werden, dass die Studierenden kurz vor dem Berufseinstieg stehen und ihre aktuelle Hauptaktivität das Studium darstellt[181]. Es kam vor, dass die Bereiche wie das Berufliche, Akademische und Soziale miteinander verknüpft wurden oder überlappend diskutiert wurden. Gesellschaftlich wird, wie auch in Stellenausschreibungen nach vollumfänglicher Persönlichkeitsentwicklung gesucht und beispielsweise mit Slogans wie ”wer dem Profil entsprechen will, muss nicht nur beruflich qualifiziert sein, sondern muss eine Reihe persönlicher Eigenschaften besitzen [...] darunter unter anderem Optimismus, richtige Einstellung, Motivation, Stressresilienz, Selbstbewusstsein, Kreativität - die Liste ist lang” (Schmelzer, 2015) geworben. In den Interviews wurde dies selten erkannt: Die wenigsten beschrieben die Arbeit an sich selbst als vollumfänglich. Dennoch darf festgehalten werden, dass die Selbstoptimierung bei einigen Personen - es sei hierbei an die neun Personen gedacht, die angaben, die beste Version von sich selbst sein zu wollen - breite Teile der Persönlichkeit betrifft. In der Studie von Mühlhausen (2016b, S. 5) ist auch davon die Rede, dass Selbstoptimierung immer mehr in die Freizeit-Bereiche eintritt und beispielsweise das Herstellen von ‘work-life-balance’ zur Selbstoptimierung gehört. Die Resultate der Interviews lassen den Schluss zu, dass die Studierenden sowohl im professionellem Kontext als auch in privaten Bereichen so beispielsweise den eigenen Beziehungen, der Alltagsgestaltung und der eigenen Persönlichkeit an sich arbeiten und sind folglich mit der bestehenden Forschung konform[182].

Soziologe Rosa (2019, S. 40) spricht von einem Ressourcenfixierungs-Trend der Selbstoptimierung: ”Gesundheit, Geld, Gemeinschaft (bzw. Beziehungen), dazu häufig noch Bildung und Anerkennung, gelten nicht nur als die wichtigsten Ressourcen für ein gutes Leben, sie werden für das gute Leben selbst gehalten”. Rosa (2019) argumentiert weiter, dass eine ‘Parametrisierung der Lebensführung’ stattgefunden hat, die es ermöglicht das Leben mit Hilfe messbarer Einheiten zu vergleichen und zu optimieren. Die Parametrisierung der Selbstoptimierung kam in den Interviews kaum zur Sprache[183]. Ein Fokus auf die Vermehrung der eigenen Ressourcen kann, entsprechend den Kapitalbegriffen von Bourdieu festgestellt werden[184]: Soziale Kontakte knüpfen, diese pflegen, das eigene Netzwerk ausbauen, das Erwerben sowie Ausweisen diverser Fähigkeiten

[180] Beim Vergleich gilt zu berücksichtigen, dass die Forschungsgruppe der genannten Studie ein grösseres Altersspektrum umfasste, sich insbesondere auf den Gesundheitsbereich fokussierte und ein anderes Forschungsdesign, inklusiver anderer Ausdrücke, angewendet wurde wie in dieser Bachelorarbeit.

[181] Die befragten Studierenden befinden sich hauptsächlich im Bachelor. Knapp die Hälfte der Befragten arbeitet neben dem Studium gegen Bezahlung. Knapp ein Drittel leistet freiwillige Arbeit.

[182] Entsprechend der explorativen Natur dieser Bachelorarbeit ist die Anzahl der existierenden Forschung limitiert. Insbesondere hinsichtlich der Arbeit in sozialen Beziehungen und den verwendeten Methoden zur Selbstoptimierung existiert keine der Autorin vorliegende Forschung.

[183] Es wurde beispielsweise nicht von Schrittzählern, Blutzuckermessgeräten, Zahlen auf den sozialen Medien oder ähnlichem gesprochen, was nicht bedeutet, dass die Befragten sich darin und damit nicht messen. Der Schluss, dass die Fragen nicht darauf abzielten, liegt näher.

[184] Vgl. Kapitel 2.2.2 Inwiefern dies für die Studierenden als Voraussetzung für ein gelingendes Leben angesehen wird, wird im nächsten Unterkapitel behandelt.

sowie akademischer Abschlüsse, und Zeit in sich selbst zu investieren sind nur einige genannte Beispiele. Es sei zudem darauf hingewiesen, dass es durchaus befragte Studierenden gab, die ihre eigenen Selbstoptimierungs- und Selbstinvestitionsnarrative hinterfragen und sich teilweise davon abzugrenzen versuchen.

Abschliessend lässt sich über die Selbstoptimierungstendenzen der Befragten sagen, dass diese nicht aus der Norm fallen. Primär wird entsprechend dem vorherrschenden Zeitgeist an sich gearbeitet. Obwohl die Tätigkeiten und Bereiche teilweise auseinandergehen, stehen die Ergebnisse nicht im Gegensatz zur Theorie und der bisher limitiert existierenden Forschung.

5.3. Selbstoptimierung und Glücksstreben

Die Selbstoptimierung wurde schon in verschiedenen Studien mit dem Streben nach Glück und dem gelingenden Leben in Verbindung gebracht (Girkinger, 2019; Mühlhausen, 2016a). Die Mehrheit der Befragten berichtet in den Interviews davon, dass sie bei der Arbeit an sich ein gelingendes Leben als Ziel vor Augen haben. Obwohl die Interviewten teilweise unterschiedliche Selbstoptimierungspraktiken und -verständnisse sowie unterschiedliche Vorstellungen von einem gelingenden Leben haben, sieht die Mehrheit der Studierenden[185] Selbstoptimierung als wegebnend für ein gelingendes Leben.

In den Interviews wird Anerkennung und das 'Dazugehören' als häufigste extrinsische Motivation genannt. Dass eine deutliche Mehrheit der befragten Personen von Zugehörigkeit und Zustimmung zur Arbeit an sich selbst motiviert ist, untermauert die Thesen von Rosa und Bourdieu hinsichtlich der Ressourcen- und Kapitalakkumulation. Vom durch die Selbstinvestition erworbenen Kapital und den Ressourcen geht Anerkennung aus[186]. Das Anerkennungsnarrativ passt wiederum mit dem zuvor beschriebenen Gesellschaftsbild sowie den als Religionsersätze diskutierten Imperativen 'an sich selbst zu arbeiten' oder 'etwas aus dem Leben zu machen' zusammen.[187]

Glück ist im Fragebogen der beliebteste Grund für die Arbeit an sich selbst[188]. Auch in der Befragung von (Mühlhausen, 2016a) wurde Glück als dritt-häufigster Grund für die Selbstoptimierung im Gesundheitsbereich genannt. In der Folge werden die Glücksvorstellungen der Studierenden eingehend diskutiert. Entsprechend der Heldenreise ist Glück stark mit sich kennen(lernen) sowie der Sinnhaftigkeit des eigenen Lebens verbunden. Im Fragebogen haben die Studierenden der Sinnhaftigkeit des eigenen Lebens, sich kennenlernen und weiterentwickeln sowie positiven Beziehungen für ein gelingendes Leben den höchsten Wert beigemessen. Auch in den Gründen, weshalb Selbstoptimierung ausgeführt wird, sind neben dem erhofften langfristigen Glück, kurzfristigem Wohlfühlen, der Sinn und sich kennenlernen unter den beliebtesten Gründen. Es bestehen Sinnbedürfnisse und Bedürfnisse sich selbst kennenzulernen und zu entwickeln. Aber wie Zukunftsforscher Robert Junk in einem Arbeitspapier argumentiert, "gibt Selbstoptimierung keine Antwort darauf, was ein gelingendes sinnvolles Leben ausmacht" (Girkinger, 2019, S. 9; Jungk, 1993). Auch Röcke (2021, S. 211) hält es zumindest für umstritten, dass Selbstoptimierung "einen einheitlichen Wertehorizont schaffen kann, der von innen heraus die Lebensgestaltung im Sinne eines einheitlichen Prinzips anzuleiten vermag".

Laut Kolmar wird in der säkularen Gesellschaft in Lebensprojekten nach Sinn, Glück und dem erfüllten Leben gesucht (Kolmar, 2021, S. 49). Mehr als die Hälfte der Befragten erhofft sich von Lebensprojekten wie einer Familie, einer erfüllenden Tätigkeit oder der Selbstoptimierung Glück. Es scheint, als seien Tiefe und Zugehörigkeit für die Interviewten primär in Partnerschaften, Freundschaften und der Familie realisierbar und als würde Arbeit an sich zum Erreichen dieser Empfindungen dazugehören[189].

Sinn wird von den Befragten abgesehen von in erfüllten Beziehungen auch in erfüllten Tätigkeiten gesucht. Person 22 setzt Berufs- und Sinnsuche gar gleich (Transkript, Absatz 78). Knapp die Hälfte der Studierenden spricht davon, mit ihren Tätigkeiten Gutes zu tun, etwas zu bewirken oder einen Unterschied zu machen und dadurch Erfüllung zu erhalten[190]. Es lässt sich vermuten, dass einige Studierende, da sie sich auf keine innere Reise begeben, sondern Sinn in Lebensprojekten suchen, einem oberflächlicheren Verständnis von Sinn hinterherlaufen[191].

Ein ähnlich oberflächliches Verständnis lässt sich hinsichtlich des Bedürfnisses[192] und der Umsetzung des 'sich selbst Kennenlernens' vermuten. Vormbusch (2016) ver-

[185] Dies geht aus den Antworten auf die Frage hervor, ob die Studierenden Zusammenhänge zwischen ihren Antworten zu Fragen zur Arbeit an sich selbst und Fragen zu ihren Vorstellungen eines gelingenden Lebens sehen. Es sei hier darauf hingewiesen, dass die Inklusion von 'Zusammenhängen' in der Frage bei den Befragten eine Erwartungshaltung ausgelöst haben könnte, Zusammenhänge zu sehen. Darauf wird in den Limitationen der Arbeit erneut unter dem Stichwort Konfabulation eingegangen.

[186] Bourdieu bezeichnet dies als symbolisches Kapital.

[187] Die in dieser Arbeit erfragten intrinsisch und extrinsischen Selbstoptimierungsmotivationen stellen ein spannendes Forschungsfeld dar, welches im Rahmen dieser Arbeit nicht umfassend behandelt werden konnte.

[188] Glück ist in allen Bereichen ausser dem Aussehen (mitunter) der beliebteste Grund für die Arbeit an sich selbst. Im Bereich der Beziehungen liegen das Glück und die Erreichung des persönlichen Lebensziels gleichauf.

[189] Die beliebtesten Motivationen, um an den sozialen Beziehungen und Fähigkeiten zu arbeiten sind das eigene Glück und das Erreichen des eigenen Lebensziels. Dem Erreichen des persönlichen Lebensziels wurde nur im Bereich Glück stärker zugestimmt. Inwiefern sich das Arbeitsverständnis in Beziehungen über die Zeit verändert hat und inwiefern dies mit fortschreitender Selbstoptimierung zusammenhängt, könnte Ausgangspunkt einer weiterführenden Studie sein.

[190] Ein Grossteil der Studierenden nennt in den Interviews den positiven Einfluss der Arbeit an sich selbst auf das Umfeld und die Umwelt als Motivation.

[191] Sinn wird lediglich sieben Mal und vielfach in Verbindung mit Arbeit oder einer positiven Wirkung auf andere genannt. Der Kode zur erfüllenden Tätigkeit fand ebenso häufig Anwendung. Es ist zudem interessant festzuhalten, dass der Sinn im nicht vorstrukturierten Interviewformat weniger Bedeutung erhält als im vorstrukturierten Fragebogen.

[192] Bei der Befragung sowie in den Interviews wurde klar, dass ein Bedürfnis existiert, sich kennenzulernen. Sich selbst kennenlernen und entwickeln,

steht Selbstoptimierung als einen Prozess des ”Sich-selbst-Entdeckens”. Im Fragebogen war ‘sich selbst kennenlernen’ im Durchschnitt einer der beliebtesten Gründe für die Selbstoptimierung der Befragten. ähnlich wie beim Sinn[193], wird für die Wichtigkeit, die dem ‘sich selbst Kennenlernen und sich Weiterentwickeln’ im Fragebogen beigemessen wird, in den Interviews wenig davon gesprochen. Im Interview drehten sich die Narrative des sich Kennenlernens vielmehr darum, mit sich selbst und den eigenen Entscheidungen zufrieden beziehungsweise im Reinen zu sein. Einige Befragte sprechen davon, möglichst viel kennengelernt haben zu wollen, um herausfinden zu können, was ihnen gefällt. Entsprechend diesem Verständnis findet das sich Kennenlernen über das Aneinanderreihen von Erlebnissen statt. Vergleichbar beschreibt Rosa die Idealvorstellung eines modernen Lebens als eines, in dem man viel erlebt. Rosa (2005) führt aus, dass eine Beschleunigung unserer Leben notwendig ist, um möglichst viel von der Welt zu erleben. Rosa’s These erinnert auch an die Tendenz einer Vielzahl von Studierenden, die sich selbst mit neuen Eindrücken, Perspektiven und Aufgaben zu betäuben. Mithilfe der vielfach genannten Selbstreflexion scheint versucht zu werden, das Erlebte einzuordnen. Auch hier lässt sich die Frage aufwerfen, ob die Selbstreflexion auf einer tieferen Erfahrungsebene ansetzt oder verkopft und oberflächlich stattfindet. Obschon wenige der Studierenden auf die Frage nach dem gelingenden Leben innere Heldenreisen-ähnliche sich kennenlernen Narrative mitteilen, kann die These gewagt werden, dass die Mehrheit der Befragten sich nicht auf die tiefsten Ebenen des sich Kennenlernens einlässt, da sich die wenigsten ihren tiefsten Ängsten und Dämonen stellen.[194]

Es ist plausibel anzunehmen, dass das unvollständige Verständnis von Sinn und Selbstkenntnis der Studierenden zum Gefühl führen kann, um eine unerreichbare Mitte zu streben, wie es Lacan als natürlichen Teil der modernen menschlichen Existenz beschreibt (Lacan, 1977, 1992). Dieses ist in den Interviews regelmässig aufgetreten[195]. Lacan bezeichnet ‘das Ding’ auch als Inschrift eines Unverständnisses dessen, was das Selbst ist (Kolmar, 2021, S. 278 f.). Obschon eine Ungewissheit besteht, ob tiefergreifendere Verständnisse der eigenen Existenz und Sinns in den Interviews nicht eingehend genug erfragt wurden[196], erlauben die bisher diskutierten Resultate folgende Antwort auf die Forschungsfrage dieser Arbeit:

Wenn, basierend auf oberflächlichen Verständnissen von Glück, Selbstkenntnis und Sinn, in äusseren Heldenreisen wie beispielsweise der Selbstoptimierung nach eben diesen Erfahrungen[197] gesucht wird, sind diese Glückserreichungsversuche erstens zum Scheitern verurteilt und verunmöglichen zweitens möglicherweise innere Heldenreisen.

Andere Studien kommen auf ähnliche Thesen - zumindest, was den Glücksbegriff betrifft: Es sei gefährlich, wenn gesellschaftlich suggeriert wird, dass Glück durch Diäten, Konsum, Alltagsgestaltung und die Arbeit an sich selbst erreichbar ist (Girkinger, 2019, S. 7). Die Individuen bekommen zu verstehen, dass Glück ”nur eine Frage der Einstellung und Lernwilligkeit” (Girkinger, 2019, S. 20) ist. Die Ratgeberkultur lässt die Bevölkerung durch ihre unzähligen Anleitungen und Aufforderungen glauben, dass immer mehr Menschen ein erfülltes Leben führen, während das Stresslevel und die Anzahl psychischer Krankheiten in der Bevölkerung zunehmen (Perriard & Ramaciotti, 2003).

Nach dem Wissen der Autorin existieren keine vergleichbaren Studien zu den Verständnissen von Sinn und Selbstkenntnis. Hier sei nochmals auf die Theorie zurückgegriffen: Wie C. Taylor (2007) und Kolmar (2021) argumentieren verschwinden die Bedürfnisse nach Sinn, Tiefe und Zugehörigkeit, die durch innere Heldenreisen befriedigt werden können, nicht. Taylor spricht davon, dass sich durch die unbefriedigten Bedürfnisse Spannungen aufbauen, die sich dann unkontrollierbar in alle Richtungen entladen können. Aufgrund des begrenzten Rahmens dieser Arbeit wurde nicht auf diese Nova-Effekte eingegangen[198].

Der Umgang mit potenziell oberflächlichen Verständnissen von Sinn und Selbstkenntnis soll in der Arbeit nicht ausgeklammert werden, besonders weil diese Unverständnisse innere Heldenreisen verunmöglichen können: Möglichkeiten und Unmöglichkeiten von Heldenreisen im gesellschaftlichen Kontext, insbesondere welchen Raum und Diskurs wir Heldenreisen widmen, soll im nächsten Kapitel beleuchtet werden. Das nachfolgende Kapitel ordnet die Ergebnisse dieser Arbeit zudem in einem grösseren Kontext ein und liefert in Form eines Ausblicks einen Nährboden für weitere Studien, auf die konkret nach der Erläuterung der Limitationen und Potentiale dieser Arbeit eingegangen wird.

5.4. Gesellschaftlicher Raum für Heldenreisen

Mehr als die Hälfte der Entwicklungsgeschichten der Studierenden lässt sich in transformative Heldenreisen einteilen: Ein kleiner Teil davon befindet sich auf inneren Heldenreisen. Es ist denkbar, dass die Heldenreisen der Studierenden nicht spezifisch genau erfragt werden konnten oder von den Studierenden zu undeutlich formuliert wur-

ein essenzieller Teil der Heldenreise wurde in der Frage zum gelingenden Leben gleich hoch bewertet, wie alle anderen PERMA-Kategorien. Sich selbst kennenlernen geht auch als einer der vorherrschendsten Gründe um an sich selbst zu arbeiten hervor. Es ist der viertbeliebteste Grund über alle Bereiche zusammen und erhält sechsmal den höchsten Zustimmungswert.

[193] Vgl. vorletzte Fussnote.

[194] Es wird dazu geraten, die Thesen zum oberflächlichen Verständnis von Sinn und Selbstkenntnis in weiterführenden Studien zu untersuchen.

[195] Ein weiteres Beispiel: Die Mehrheit der Personen, die den Kreis als Darstellung für ein gelingendes Leben wählten, bezeichneten das Leben zudem als Kreislauf, beschrieben dessen Verlauf oder Inhalt jedoch nicht weiter.

[196] Dies weil die Interviewerin sich einerseits an den Interviewleitfaden hielt und andererseits sprachliche Grenzen bestehen, transformative Erfahrungen zu beschreiben.

[197] Gemeint sind Glück, tiefe Selbstkenntnis, Verbundenheit und Sinn.

[198] Zusätzliche Umfragen zu potenziellen gesellschaftlichen Spannungen sowie deren Entladungsmöglichkeiten sind von weiterem Interesse.

den, da diese nicht in Heldenreisen-ähnlichen Dimensionen über ihr Leben nachdenken[199].

Lacan argumentiert hingegen, dass unsere Kultur und unser Zugehörigkeitsgefühl in der Krise ist und wir uns eingestehen sollten: "Unsere Vorstellung, dass die Menschen schon gesättigt sind, wenn sie in Shoppingmalls einkaufen und im Cafe sitzen können oder eine sichere 40-Stunden-Woche haben, ist zum Irrglauben geworden" (Medick, 2015). Lacan beschreibt die modernen Menschen als suchend: "Viele Menschen wollen auch gerne für etwas kämpfen. Wer sich einmal in den Pariser Banlieues mit Jugendlichen unterhalten hat, wird das bestätigen können. Viele von ihnen hadern mit ihrer Identität und ihrer Rolle" (Medick, 2015). Die Zitate suggerieren, dass unsere Kultur zu wenig Angebote für innere Heldenreisen macht und dadurch extreme äussere Heldenreisen an Attraktivität gewinnen.

Auch Ullrich (2016) argumentiert, dass in unserem gesellschaftlichen Diskurs, ausgedrückt beispielsweise in Werbungen "kein Platz für tragische Helden, existenzielle Opfer, komplizierte Konflikte ist. Nie wird eine Stimmung vielschichtig, ein Motiv unheimlich, ein Topos grausam. Vielmehr gibt es nur strahlende Sieger, Erfolgsgeschichten, 'happy ends', glückliche und konfliktfreie Beziehungen. Und wenn doch einmal ein Gefühl jenseits guter Laune angesprochen ist, dann sicher so gut in Kitsch verpackt, dass einmal mehr alle Komplikationen verschwinden" (Ullrich, 2016, S. 197).

Ein möglicher Umgang mit den Ergebnissen dieser Studie bestünde darin, den Diskurs zum Thema Heldentum anzuregen, gesellschaftlich mehr Verständnis und Bewusstsein für die eigenen Bedürfnisse, beispielsweise nach Sinn und Zugehörigkeit, zu schaffen. Welche konkreten Massnahmen die Studierenden dabei unterstützen würden, von einer äusseren auf eine innere Heldenreise zu wechseln, könnte eine Forschungsfrage für weitere Studien darstellen. Da ansatzweise mehr Befragte, die den Kurs Ökonomie des Glücks besucht haben, eine innere Heldenreise erleben, sollten Unterrichtsfächer[200], die Glücksnarrative hinterfragen, im Detail auf ihre Wirkung geprüft werden.[201]

Eine weitere mögliche Konsequenz wären Korrekturen im Glücksdiskurs. Taylor postuliert, dass aufgrund der Säkularität der Welt keine inneren Heldenreisen gelingen können (Kolmar, 2020, S. 41). Die Beurteilung, ob innere Heldenreisen in der säkularen Gesellschaft gelingen können oder nicht, liegt ausserhalb des Rahmens dieser Arbeit. In dieser Arbeit wurden Tendenzen festgestellt, dass Glücksversprechen von Lebensprojekten, aufgrund von fehlendem Verständnis, innere Heldenreisen verunmöglichen. Kolmar (2021) stimmt insofern zu, dass sich die Individuen

> "zu schnell auf die Richtigkeit der nächsten Stufe der Erkenntnis von Wirklichkeit [verlassen] und daraus weitreichende Schlussfolgerungen für das eigene Leben und die Gesellschaft [ziehen]. All diese Theorien sind aus buddhistischer Sicht mehr oder weniger hilflose Schnellschüsse, die höchstens zufällig mal ins Ziel treffen, weil man innerhalb des westlichen Denkens die Reise nach innen nicht nur konsequent nicht antritt, sondern eine solche auch noch regelmässig als unwissenschaftlichen Mystizismus disqualifiziert und Menschen belächelt, die eine solche mache" (S. 279).

Thema zukünftiger Studien könnte sein, wie wir Selbst- und Menschenkenntnis diskutieren und insbesondere innere Heldenreisen vermehrt in den öffentlichen Diskurs einbringen können. Um das Konkurrenzpotential zwischen inneren und äusseren Heldenreisen zu verringern, müsste gesellschaftlich am Glücksverständnis gearbeitet werden. Es wäre sinnvoll, öffentlich das Versprechen "der Herstellbarkeit von Glück und Erfolg zu hinterfragen" (Girkinger, 2019, S. 64) und realistische Erwartungen sowie fundierte Methoden hinsichtlich der Erfahrung von Sinn zu liefern.

Es könnte zukünftig angedacht werden, einen informativ aufklärenden statt ratgebenden Diskurs zu führen. Die Vielseitigkeit von Glück sollte zum Ausdruck kommen. Der bisherige Diskurs der Selbstverantwortung, Machbarkeit und Multioptionalität (Girkinger, 2019, S. 37 ff.) hat oberflächlich positive Seiten, kann aber auch zur Überforderung des Individuums führen. Es ist darauf zu achten, dass der Diskurs nicht zwangspositiv ist. Reckwitz findet dafür folgende Worte: "Die Spätmoderne ist im Kern eine Kultur der positiven Affekte, die den negativen oder auch nur ambivalenten Erfahrungen kaum legitimen Raum gibt" (Reckwitz, 2017, S. 347). Das Glück ist aber eben durch Lebensprojekte erreichbar: "Wir können uns um unser Glück und um das, was wir als wertvoll erachten kümmern und bemühen, aber nicht mit Checklisten herstellen, genauso wenig wie wir Liebe oder Freundschaft herstellen können" (Girkinger, 2019, S. 66).

Dies legt die Empfehlung nahe, dass es ambivalente Erfahrungen aushalten zu lernen gilt. Weil sich viele Dinge in der säkularen Welt nicht erreichen oder ändern lassen, ist gesellschaftlich und individuell empfohlen zu "lernen mit der Leere umgehen zu können" (Kolmar, 2021, S. 177). "Gerade das Loslassen, das Sich-Abfinden, das Akzeptieren von Grenzen kann der Schlüssel sein, um gut mit dem leben zu können, was wir haben" (Girkinger, 2019, S. 67). Das sind Beispiele von inneren Heldennarrativen, die zentralen wozu-Fragen nicht durch wie-Fragen ersetzt, sondern uns durch die Konfrontation mit unseren inneren Dämonen Zugehörigkeit und Sinn empfinden lässt. Der Rocksänger Bruce Springsteen

[199] Basierend auf der Annahme, dass mehrere Heldenreisen aneinandergereiht in einem Leben gemacht werden können und somit je nach Lebensabschnitt unterschiedliche Heldenreisen möglich sind, wäre es von Interesse in weiterführenden Studien die potenziellen inneren und äusseren Heldenreisen von Bachelor und Master Studierenden zu untersuchen. Der Auszug von zu Hause, möglicherweise sogar aus dem Heimatland könnten solche Heldenreisen darstellen. Eine Befragung mit Fokus auf persönliche Gedanken, Ziele sowie Bedürfnisse (entsprechend der Maslow Pyramide) der Studierenden könnte dafür aufschlussreich sein.

[200] Ebenso könnte persönliches Mentoring als Massnahme geprüft werden.

[201] Es existieren beispielsweise Studien, die belegen, dass das Besuchen von Stärken-fokussierten Positiven Psychologie Kursen sich positiv auf das Wohlbefinden von Studierenden auswirkt (B. W. Smith et al., 2021).

ist davon überzeugt, dass ”mit jedem Jahr, das vorbei geht, der Preis der Verweigerung sich seinen inneren Dämonen zu stellen steigt” (Hainey, 2018).

Transformative Heldenreisen sind zentrale Pfeiler für das Wohlergehen und die psychische Gesundheit einer Gesellschaft (Kolmar, 2021, S. 95). Nachdem in dieser Arbeit Tendenzen festgestellt wurden, dass oberflächliche Verständnisse der Erreichbarkeit von Sinn und Glück durch Lebensprojekte wie die Selbstoptimierung existieren, stellt dieses Kapitel insbesondere Ansatzpunkte dar, weitere Studien zu gesellschaftlichen Selbstoptimierungs- und Glücksnarrativen zu lancieren und basierend darauf eine kritische öffentliche Debatte zum Wohle der Gesellschaft anzuregen.

Konkrete Vorschläge zu weiterführende Studien sind im separaten Kapitel 6.4 aufgeführt. Vorerst wird auf die wichtigsten Limitationen dieser Arbeit eingegangen.

6. Limitationen und Potenziale

6.1. Forschungsdesign und allgemeine Limitationen

Die grösste Limitation dieser Arbeit besteht in der geringen Zahl der Teilnehmenden und der daraus folgenden limitierten Aussagekraft der Resultate[202]. Bei explorativen Studien, die mit einer geringen Zahl Personen das Forschungspotential eines Themas erkundet, liegt das vielfach in der Natur der Sache. In dieser Arbeit war es lediglich mit dieser Anzahl Teilnehmenden möglich, die aufwendige Kombination quantitativer und qualitativer Forschungsmethoden anzuwenden.

Die Methodenkombination wurde angewendet, um ein vielfältigeres Bild der Selbstoptimierungsnarrative zu erhalten und eine Scheinsicherheit der Daten abwenden zu können. Eine Problematik davon besteht darin, dass die erhaltenen Daten widersprüchlich sein können. In dieser Arbeit sind die Daten der Interviews mit jenen des Fragebogens kompatibel. Es ist möglich, dass die Interviews das anschliessende Ausfüllen des Fragebogens beeinflusst haben und die hohe Kongruenz der Daten unter anderem so entstanden ist. Allerdings wurden die Testinterviewten in den Pretests hinsichtlich ihrer wahrgenommenen Beeinflussung des Interviews auf den Fragebogen befragt: Sie gaben an, dass sie die Abfolge von Interview und Fragebogen gelungen finden und sich dadurch beim Ausfüllen des Fragebogens unterstützt fühlten, diesen aber nicht bewusst anders ausfüllten. Obwohl eine Beeinflussung nicht ausgeschlossen ist, schien die Kombination der Methoden das bestmögliche Forschungsdesign für diese explorative Studie.

Im Verlauf dieses Kapitels wird auf Fragebogen- und Interview-spezifische Limitationen eingegangen. Vorerst werden weitere allgemeine Limitationen behandelt. Aufgrund des hohen zeitlichen Aufwands für die Teilnehmenden wurden diese unter anderem aus dem erweiterten Bekanntenkreis rekrutiert. Da der Interviewerin zudem hauptsächlich Zugriff auf Kommunikationskanäle der Universität St. Gallen hatte, studieren die meisten befragten Studierenden dort. Es ist möglich, dass der Leistungsgedanke an der führenden Schweizer Universität für Wirtschaftswissenschaften besonders ausgeprägt ist.[203] Ein weiteres Selektion Bias ergab sich hinsichtlich an Nachhaltigkeit interessierter Studierenden. In weiteren Studien könnten durch finanzielle Mittel Anreize zur Teilnahme geschaffen werden, entsprechend grossflächiger rekrutiert werden und Aussagen über verschiedene Universitäten oder Ausbildungswege hinweg gemacht werden.

Weitere Limitation des Forschungsdesign liegen in der Subjektivität[204] und der Abhängigkeit von der Tagesform[205] der befragten Studierenden. Die Studierenden wurden jeweils nur an einem Tag befragt und hatten keinerlei Vorbereitungszeit. Das Forschungsdesign konnte aufgrund des limitierten Rahmens dieser Arbeit das aktuelle Wohlbefinden der Befragten nicht miteinbeziehen.[206]

Des Weiteren konnte den potenziell unterschiedlichen Ausgangspunkten der Befragten in dieser Arbeit begrenzt Rechnung getragen werden. In anderen Worten wurde im Forschungsdesign limitiert erfasst, wie intensiv sich die Befragten bereits mit Vorstellungen eines gelingenden Lebens sowie der Arbeit an sich selbst befasst haben. Obwohl selbstreflexive Fragen im Fragebogen grob abbilden sollten, wie stark sich die Befragten mit den Themen Selbstoptimierung und dem gelingenden Leben auseinandergesetzt haben, kann dies für die Interviews nicht gesagt werden. Es konnte nicht ausgeschlossen werden, dass Personen unterschiedlich viel über Selbstoptimierung reflektiert haben. Es wurde keine Kontrollfrage zur Messung des Masses an bisheriger Selbstreflexion eingefügt, weshalb keine Aussagen zur Reflexivität der Studierenden in dem Bereich getroffen werden können[207] und sich dies limitierend auf die Aussagekraft der Resultate auswirkt.

Es wurde zudem in den Interviews nicht konkret nach den Verständnissen von ‘an sich Arbeiten’ gefragt. Dies wurde be-

[202] Entsprechend stark können einzelne Geschichten und Schicksale ins Gewicht fallen und die Resultate in eine Richtung beeinflussen. Eine befragte Person verlor kurz vor dem Interview eine geliebte Person, bei einer anderen handelt es sich um eine Person mit psychischen Erkrankungen. Die Themen wurden in den Interviews angesprochen, ohne dass ein Auswertungsgefäss dafür existierte. In künftigen Studien könnten insbesondere psychischen Krankheiten im Interview aktiv aufgegriffen werden. Denn ”was wir für eine psychische Krankheit halten—bzw. was von manchen Therapeuten sodefiniert wird—ist in Wirklichkeit ein Transformationsprozess” de (Kolmar, 2021, S. 208 ff.).

[203] Im Anhang C können die Universitäts-spezifischen Narrative eingesehen werden.

[204] Dazu ein Beispiel: Was für Person 2 ‘stimme voll und ganz zu’ bedeutet, kann für Person 1 unter stimme zu fallen.

[205] Dass die Tagesform der befragten Personen in qualitativen Studien die Resultate beeinflussen kann, zeigt unter anderem Andreas Diekmann (Diekmann, 2007, S. 463)

[206] In künftigen Studien wäre es wünschenswert, zumindest eine Einschätzung des aktuellen Wohlbefindens zu erfassen, wenngleich dieses erneut subjektiv ist. Anstelle von Vorstellungen des gelingenden Lebens könnte mittels des originalen PERMA-Profilers die aktuelle subjektive Lebenszufriedenheit mit den jeweiligen Selbstoptimierungsnarrativen und -praktiken untersucht werden.

[207] Es wird hier angenommen, dass die Resultate je nach Level der Reflektiertheit unterschiedlich eingeordnet werden können.

absichtigt, um die spontanen Reaktionen und Verständnisse der Studierenden einzufangen und durch eine Definition des Begriffes vor den Selbstoptimierungsfragen im Fragebogen ausgeglichen. Für die Interviews gab es keine spezifische Verständnisfrage zum Begriff an sich Arbeiten und neben dem Kodierleitfaden kein Auswertungsverfahren der Verständnisse. Obwohl in den Pretests das Verständnis von an sich Arbeiten konkret erfragt wurde und dies dem gewünschten Verständnis entsprach, besteht eine Limitation der Arbeit darin, dass Studierende unter dem Begriff an sich Arbeiten unterschiedliche Situation verstehen[208]. Aufgrund der bedeutenden Rolle des 'an sich Arbeitens' Begriffs für diese Arbeit wurden die unterschiedlichen Auffassungen des Begriffs unter 4.2 in den Resultaten erfasst.

6.2. Interview-spezifische Limitationen

Die Interviewerin dieser Arbeit schulte sich durch die Pretests für die Interviews. Aufgrund der fehlenden professionellen Ausbildung ist zumindest fraglich, ob die Interviewerin immer in der Lage war, genügend tief nachzufragen, wenn es aufgrund von unzureichender Beantwortung der Fragen notwendig gewesen war. Zum Zeitpunkt der Interviews war die Interviewerin zudem 23 Jahre alt und hatte damit selbst limitierte Erfahrung hinsichtlich eigener transformativer Reisen. Dies machte es zusätzlich anspruchsvoll in Echtzeit mit den richtigen Nachfragen aufzukommen.

Der begrenzte transformative Erfahrungsschatz der Interviewerin und die nicht kommunizierbare Natur transformativer Erfahrungen machten es zudem zur Herausforderung, transformative Erfahrungen anderer einzuordnen. In Form von im Kodierfaden definierten Indikatorenkodes wurden bestmögliche Voraussetzungen für eine objektive Auswertung geschaffen. Da die Einteilung in die Heldenreisenkategorien nicht immer eindeutig erschien, wurde zu jeder Einteilung zur Nachvollziehbarkeit eine Begründung geschrieben.[209]

Darüber hinaus, war es nicht ideal, dass die Interviewerin die Interviews und die Auswertung selbst vornimmt, da sie aufgrund der kleinen Anzahl Befragten und dem geringen zeitlichen Abstand zwischen Interviewdurchführung und -auswertung oftmals die Aussagen den Personen noch zuordnen konnte. Trotz der vorgenommenen Anonymisierung ist es möglich, dass somit implizite 'biases' der Befragerin[210] in die Auswertung haben einfliessen können. Mit der getreuen Verwendung des erstellten Kodierleitfadens konnten diese hoffentlich weitestgehend minimiert werden.

Eine weitere Limitation der Interviewergebnisse besteht darin, dass die Aussagen nicht mit den Befragten nachbesprochen wurden, beziehungsweise diese aufgrund des Rahmens der Arbeit nicht die Möglichkeit erhielten die eigenen Interviewtranskripte einzusehen und allfällige Missverständnise aufzuklären oder Ergänzungen vorzunehmen[211].

In den Interviews wurden zudem gesellschaftlich heikle und sehr persönliche Fragen gestellt. "Der menschliche Drang zur Selbstachtung und Selbstdarstellung kann ausserdem zu übertreiben bzw. Untertreiben, Verdrängen, Ausweichen, Lügen und zum Anpassen der Aussagen führen" ("Response-Bias: Verzerrung von Ergebnissen durch Teilnehmer - Wirtschaftspsychologische Gesellschaft", n.d.). Das Bild des weniger intensiven an sich Arbeitens im Fragebogen liesse sich auch dadurch begründen, dass die Befragten im Fragebogen aufgrund der Anonymität nicht mit potenziellen Erwartungen der Befragerin konfrontiert sind. Es wurde ein möglichst sicherer und kooperativer Rahmen geschaffen, um das Risiko Wunscherzählungen statt die Wirklichkeit von den Studierenden zu hören, minimal klein zu halten.[212]

Es wurde insbesondere versucht, nicht erwartend zu erscheinen, da Personen zu Konfabulation tendieren, wenn sie sich unter einem Antwortenzwang befinden. Bei Konfabulation handelt es sich "um eine Abwesenheit von Zweifel dort, wo Zweifel angebracht wäre" (Kolmar, 2021, S. 80). Obschon keine reine Konfabulation festgestellt werden konnte, gab es möglicherweise Sequenzen, in denen die Studierenden mehr von ihren Wunschvorstellungen als der Realität erzählt haben. Ein weiteres Risiko besteht darin, dass Personen dazu tendieren kongruente Geschichten erzählen zu wollen. In den Interviews wurde vermehrt festgestellt, dass die Studierenden sich eine kongruente Geschichte[213] auslegen. Da im Fragebogen keine persönliche Konfrontation mit der Interviewerin stattfindet, konnten mittels dessen allfällige Verzerrungen ausgeglichen werden.[214]

Da die Interviewerin weiblich ist, kann zudem mit einem 'gender bias'[215] gerechnet werden: Es ist möglich, dass Frauen offener zu einer Frau sprechen als zu einem Mann. Zudem besteht die Möglichkeit, dass Männer weniger authentisch mit der Interviewerin gesprochen haben, als sie dies gegenüber einem Interviewer getan hätten.

[208] Es sei erneut darauf hingewiesen, dass auf Begriffen wie der Selbstoptimierung weitaus stärkere Bewertungen lasten und der Begriff an sich Arbeiten somit die geringstmögliche Problematik mit sich brachte.

[209] Die Begründungen können unter folgendem Link eingesehen werden.

[210] Aufgrund des begrenzten Rahmens dieser Arbeit kann nicht auf alle potenziellen 'biases' der Interviewerin eingegangen werden. Für weiterführende Literatur sei hier insbesondere auf folgende zwei Studien verwiesen: (Kvale, 1994; Salazar, 1990).

[211] Dies könnte in weiterführenden Studien verbessert werden.

[212] Es wird davon ausgegangen, dass die geringere Intensität des an sich Arbeitens auch durch die im Fragebogen enthaltene Definition des an sich Arbeitens, die Verwendung einer Intensitätsskala sowie möglicherweise unpassende Tätigkeitsbeispiele zustande kam.

[213] Dies zum Beispiel in dem die Studierenden am Ende des Interviews einen Bogen zur Antwort der 'Warm-up'-Frage machen und damit das Interview abschliessen.

[214] In diesem Kapitel wurden die für am bedeutendsten gehaltenen Fehlerquellen dieses Interviews beschrieben. Eine Übersicht von möglichen weiteren Fehlerquellen in Interviews und weiterführende Literatur hierzu kann dem Buch von Diekmann (2007) auf den Seiten 446-471 entnommen werden.

[215] Weitere Verzerrungen können aufgrund von anderen sozio-ökonomische Faktoren, die bei der befragten und befragenden Person entweder ähnlich oder unterschiedlich sind, entstehen. Aufgrund des begrenzten Rahmens dieser Arbeit wird darauf nicht weiter eingegangen. Insbesondere zu 'gender biases' sei auf die Studie von verwiesen Williams and Heikes (1993) verwiesen.

6.3. Fragebogen-spezifische Limitationen

In der Methodik wurde bereits beschrieben, weshalb ein Mittelwert im Fragebogendesign hinzugefügt wurde. Dass die Befragten etwas mehr zur Mitte tendieren, ist eine natürliche Folge davon, schränkt jedoch je nach dem die Aussagekraft der Resultate ein. Es ist zudem plausibel, dass durch die Wiederholung des Frageformats im Fragebogen ein Trainingseffekt stattgefunden hat und die ersten Frageblöcke von den Befragten möglicherweise langsamer und konzentrierter ausgefüllt wurden.

Das Risiko, dass die Studierenden neben dem Ausfüllen des Fragebogens durch Nachrichten oder andere Aktivitäten abgelenkt werden und entsprechend unkonzentriert den Fragebogen ausfüllen, konnte durch die Anwesenheit wahrscheinlich verringert, nicht aber ausgeschlossen werden. Für künftige Studien empfiehlt es sich dennoch, eine Frage zur Kontrolle der Aufmerksamkeit oder weitere Ausschliessbarkeitskriterien für den Fragebogen zu formulieren.

Abschliessend sei erwähnt, dass die Kombination der komplexen und transdisziplinären Themen[216] Selbstoptimierung, transformative Heldenreisen und gelingendes Leben für die Autorin eine Herausforderung darstellte. Obwohl so eine Blindheit auf dem eigenen Gebiet weitgehendst ausgeschlossen werden kann, haben sich höchstwahrscheinlich weitere Verzerrungen ergeben, welche die Autorin aufgrund ihrer begrenzten psychoanalytischen Kenntnisse nicht einordnen kann. Dass sich die Befragerin zudem in ähnlichen Lebenssituationen wie die Befragten befindet, kann einerseits zu weiteren 'Biases' geführt haben. Andererseits ermöglichten die Gemeinsamkeiten ein persönliches und intimes Format, auf welchem in weiteren Studien aufgebaut werden kann.

Nach der Nennung der stärksten Limitationen dieser Arbeit sollen auf diese und weitere anhand von Verbesserungsvorschlägen eingegangen werden und die aus der Arbeit hervorgehenden weiterführenden Forschungsansätze zusammengefasst werden.

6.4. Verbesserungsvorschläge und weiterführende Studien

Ein übergeordnetes Ziel weiterführender Studien könnte die Prüfung der nachfolgenden Thesen sein, die in dieser Arbeit aufgestellt wurden; dass Studierende der Wirtschaftswissenschaften oberflächliche Sinn- und Selbstkenntnis-Verständnisse haben und diese unter gewissen Umständen[217] ihr Glücksstreben und -erfahren beeinflusst. Von weiterem Interesse wäre zudem, die vorhandenen Sinn- und Selbstkenntnis-Verständnisse mit den in der Gesellschaft vorhandenen Erfahrungsmöglichkeiten zu analysieren. Der Fokus dieser Arbeit lag auf den Selbstoptimierungsverständnissen, weshalb weiterführende Studien sich sowohl der tiefgründigeren Untersuchung derer sowie deren Konsequenzen widmen können. Basierend auf den Thesen, dass unsere Kultur zu wenig Angebote (Lacan, 1977, 1992; C. Taylor, 2007) für Heldenreisen und das Erfahren von Sinn liefert, wäre es wichtig in Erfahrung zu bringen, ob dies die Studierenden zunehmend frustriert und sie nach Wegen suchen, diese Spannungen zu entladen[218].

Eine Wiederholung der Studie mit einer erhöhten Anzahl befragter Personen könnte die Aussagekraft der Resultate stärken - vorausgesetzt diese wären vergleichbar. Darüber hinaus könnte mit mehr Teilnehmenden in zusätzliche Richtungen explorativ geforscht werden. Konkret wäre es in weiterführenden Studien denkbar, potenzielle Unterschiede hinsichtlich der Selbstoptimierung entsprechend verschiedenen Ländern[219], sozialen Klassen, dem Glauben, dem Geschlecht sowie der Ausbildung[220] zu untersuchen. Nachtwey (2016) stellte beispielsweise fest, dass sich Personen aus der mittleren sozialen Klasse am stärksten selbstoptimieren, da sich diese um ihren sozialen Abstieg fürchten.

Der Einbezug des aktuellen Wohlbefindens[221] ins bisherige Forschungsdesign könnte eine weitere erkenntnisbringende Studie hervorbringen. Inwiefern Personen auf inneren oder äusseren Heldenreisen aktuell zufriedener sind, wäre eine mögliche Forschungsfrage. Zumindest dem Fragebogen könnten Fragen zur subjektiven Lebenszufriedenheit hinzugefügt und potenzielle Zusammenhänge untersucht werden.

Allgemein konnte auf Grund des begrenzten Rahmens dieser Bachelorarbeit nicht das ganze Potential aus den gesammelten Daten geholt werden. Es wurde vielfach mit Mittelwerten gearbeitet. Es besteht weiterhin viel Potential in der Untersuchung individueller Interviewaussagen in der Kombination mit den im Fragebogen gewählten Antworten. Insbesondere mit anderen Methoden[222] könnten weitere quantitative Zusammenhänge basierend auf den Fragebogendaten erforscht werden.

Die Selbstoptimierungstätigkeiten wurden in dieser Bachelorarbeit anhand ihrer Durchführungsintensität erfragt. Eine andere Möglichkeit die Intensität zu messen wäre der persönliche Konsum, der in der Arbeit kaum zur Sprache kam. Weitere Studien zur Selbstoptimierung anhand des Konsums sind insofern von Interesse, als der Konsum der Befragten absolut besser verglichen werden kann. Zudem

[216] In welchen die Autorin, aufgrund ihrer Wirtschafts-orientierten Ausbildung, insbesondere hinsichtlich der zu verwendenden sozialpsychologischen Methoden, kaum Vorkenntnisse besass.

[217] In dieser Arbeit wurde festgestellt, dass Lebensprojekte wie die Selbstoptimierung und andere äussere Heldenreisen, unter anderem glücksversprechend wirken. Das Forschungsdesign künftiger Studien könnte den Fokus auf der Selbstoptimierung behalten. Es wäre ebenso spannend, das Forschungsdesign mit Fokus auf die Sinn- und Selbstkenntnisnarrative anzupassen.

[218] Was Taylor als sogenannten Nova-Effekt beschreibt.

[219] Hier wäre es besonders interessant Vergleiche zwischen einem stark säkulareren und weniger säkularen Land zu ziehen.

[220] Einerseits könnten verschiedene Berufs- und Ausbildungswege sowie andererseits verschiedene Universitäten in den Fokus genommen werden.

[221] Dies einerseits, weil sich die Selbstoptimierungspraktiken und -narrative bereits auf das aktuelle Wohlbefinden auswirken können und andererseits die psychische Gesundheit junger Personen und Studierender insbesondere zunehmend schlechtere Werte verzeichnet. (Barrense-Dias et al., 2021; Bundesamt für Statistik, 2018, S. 12)

[222] Hier sind insbesondere quantitative Methoden gemeint, die Aussagen zur statistischen Relevanz ermöglichen. Ein Beispiel stellen Regressionsanalysen dar. Eventuell müsste dafür das Forschungsdesign leicht angepasst werden und/oder mehr Personen befragt werden.

könnten Glücksversprechen spezifischer durch Konsum untersucht werden.

Wie bereits in anderen Kapiteln der Diskussion angetönt, könnten weitere Studien zu vorherrschenden inneren und äusseren Heldenreisen der heutigen Studierenden sowie zum gesellschaftlichen Umgang mit Heldenreisen wichtige Erkenntnisse liefern. Es wäre erstens wissenswert, welche Momente für junge Erwachsene Krisenmomente darstellen, von welchen eine innere Heldenreise gestartet werden könnte, sowie welche äusseren Heldenreisen in diesem Alter angetreten werden können. Das Zugrundegehen der ersten romantischen Beziehung, Nichtbestehen universitärer Prüfungen, ein Studienabbruch oder -wechsel sowie der Auszug von zu Hause könnten als solche geprüft werden.

Weiterführende theoretische Arbeiten könnten die Ergebnisse dieser Arbeit in einem grösseren normativen Rahmen wie dem Buddhismus oder der Tugendethik einordnen sowie konkrete Vorschläge zur Verwendung der Begriffe Heldenreise, Glück und Sinn liefern.

7. Fazit

Die Säkularisierung der (westlichen) Welt und die notwendige Anpassung der Individuen an die darin vorherrschenden gesellschaftlichen Konventionen führen laut (Kolmar, 2021; Lacan, 1977; C. Taylor, 2007) dazu, dass den Menschen im Kern ihrer Existenz etwas fehlt. Das Gefühl der Leere entspringt der unzureichenden Bedürfnisbefriedigung nach Zugehörigkeit, Bedeutung und Sinn. Während der epistemische Rahmen ungenügende Angebote zur Befriedigung jener Bedürfnisse liefert, wird durch unzählige Ratgeber, Aufforderungen und die Multioptionalität suggeriert, dass das Glück erreichbar ist, ja sogar von jedem Individuum persönlich erreicht werden muss. Transformative Entwicklungsgeschichten wie die Heldenreise nach Campbell (1953) sind in der Lage jene Bedürfnisse nach Transzendenz, Zugehörigkeit und Tiefe zu befriedigen, sind aber in der postheroischen Ordnung problematisch geworden.

Im proklamierten Zeitalter der Selbstoptimierung besteht indessen die Gefahr, dass Selbstoptimierung als glücksversprechende Entwicklungsgeschichte (miss)verstanden wird. Ziel dieser Arbeit war es, die Forschungslücke, die hinsichtlich der Verständnisse von Selbstoptimierung als Entwicklungsgeschichte und Glücksversprechen besteht, mit einer ersten explorativen, empirischen Umfrage zu schliessen zu beginnen. Diese Arbeit bezog sich zudem auf Studierende, da diese in Studien zur mentalen Gesundheit verglichen mit Gleichaltrigen sowie anderen Generationen schlechter abschnitten. Folgende Forschungsfrage stand zur Beantwortung: Inwiefern konkurrieren Selbstoptimierungsnarrative und -praktiken von Studierenden der Wirtschaftswissenschaften mit Heldenreisen, die dem gelingenden Leben zuträglich sind?

In einem ersten Schritt wurden dazu die Konzepte und fundamentalen Unterschiede zwischen der Selbstoptimierung als Beispiel für äussere Heldenreisen und der inneren Heldenreisen auf Basis einer Literaturrecherche erarbeitet. In einem nächsten erlaubte es die Methodenkombination von vorstandardisierten Interviews und einem Fragebogen, das Gesellschaftsbild der Studierenden, ihre Selbstoptimierungspraktiken und -motive, Vorstellungen eines gelingenden Lebens sowie deren Zusammenhänge qualitativ und quantitativ zu erfassen.

Aus den Antworten wird deutlich, dass die Befragten die Gesellschaft säkular und postheroisch wahrnehmen. Die Befragten bewerten zudem das Leistungserwarten höher als beispielsweise die Entwicklungsmöglichkeiten in der Gesellschaft. Fast alle der Befragten erwarten von sich, ein Leben lang an sich zu arbeiten. In Duttweiler's Worten wird das "Leben als unendliche Baustelle gesehen", auf welcher "ständige Ausbesserung in diversen Bereichen" gemacht werden können (Duttweiler, 2016, S. 47). Die Teilnehmer und Teilnehmerinnen dieser Studie optimieren sich insbesondere hinsichtlich ihres beruflichen Werdegangs, ihrer Gesundheit sowie sozialen Beziehungen. Dem eigenen Leben Sinn zu geben, kurzfristiges Wohlbefinden, das Erreichen des persönlichen Lebensziels, sich kennenlernen sowie allen voran langfristiges Glück und soziale Anerkennung sind die genannten Gründe für die Arbeit an sich selbst.

Was sie unter 'Sinn' und 'sich kennenlernen' verstehen, wurde von den Interviewten selten ausgeführt und falls schon, dann auf eine oberflächliche Weise. Die endlose Arbeit an sich selbst, die Gründung einer Familie, oder einer erfüllende Tätigkeit - idealerweise mit positivem Einfluss auf die Welt - nachzugehen, wurden als sinnvoll beschrieben. 'Sich kennenlernen' bekam vielfach den 'mit sich im Reinen sein'-Stempel, was so viel bedeutete wie am Ende des Lebens nichts zu bereuen. Die wenigsten Befragten stellen sich in Form von inneren Heldenreisen ihren eigenen Dämonen und erlangen entsprechend tiefe Kenntnis über sich selbst, die Welt, und sich selbst als deren Teil. Äussere Heldenreisen werden, möglicherweise basierend auf den oberflächlichen Sinn- und Selbstkenntnis-Verständnissen, als Weg zum Glück wahrgenommen und verunmöglichen durch diesen Fehlglauben potenziell innere Heldenreisen.

Obschon diese Arbeit die Forschungsfrage somit übereinstimmend mit der bestehenden Literatur, die besagt, dass gefährliche Glücksversprechen in der Selbstoptimierung proklamiert werden und dies zur Überforderung und Erschöpfung der Individuen führen kann, beantwortet, ist weitere Forschung zur statistischen Legitimation der Resultate dieser Arbeit mit einer höheren Anzahl Teilnehmenden benötigt. Die Ergebnisse dieser ersten interdisziplinären Arbeit zu den Verständnissen von Selbstoptimierung, Vorstellungen erfüllter Leben und deren Zusammenspiel mit Heldenreisen liefern zudem Ansatzpunkte für weitere Studien zu den Verständnissen von Sinn und Selbstkenntnis sowie Zusammenhänge zwischen Selbstoptimierung und mentaler Gesundheit.

Weitere Forschung und die Beobachtung des Diskurses rund um Heldenreisen sind von zentraler Bedeutung, da transformative Heldenreisen zentrale Pfeiler für das Wohlergehen und die psychische Gesundheit einer Gesellschaft sind (Kolmar, 2021, S.95).

References

Aghamanoukjan, A., Buber, R., & Meyer, M. (2009). Qualitative Interviews. In R. Buber & H. Holzmüller (Eds.), *Qualitative Marktforschung: Konzepte–Methoden–Analyse* (pp. 415–436). Gabler.

Aichinger, H., & Bauer, K. (2014). Zwischen Beschleunigung und Abstellgleis. *DerStandard*. Retrieved May 17, 2023, from https://www.derstandard.at/story/2000009299351/45-plus-zwischen-beschleunigungswahn

Albert, M., Hurrelmann, K., Quenzel, G., & Schneekloth, U. (2016). Die 17. Shell Jugendstudie – Eine pragmatische Generation im Aufbruch. *Diskurs Kindheits- und Jugendforschung*, *11*(2), 241–246. https://doi.org/10.3224/diskurs.v11i2.23658

Allison, S. T., & Goethals, G. R. (2017). The hero's transformation. In *Handbook of heroism and heroic leadership* (pp. 401–422). Routledge.

Allison, S. T., Goethals, G. R., Marrinan, A. R., Parker, O. M., Spyrou, S. P., & Stein, M. (2019). The metamorphosis of the hero: Principles, processes, and purpose. *Frontiers in Psychology*, *10*, 1–14. https://doi.org/10.3389/fpsyg.2019.00606

Backstrom, C. H., & Hursh, G. D. (1963). *Survey Research*. Northwestern University Press.

Balandis, O., & Straub, J. (2018). Selbstoptimierung und Enhancement. *Journal für Psychologie*, *26*(1), 131–155. https://doi.org/10.30820/8247.09

Barnes, J. (2008). Nothing to Be Frightened of. *The New York Times*. Retrieved May 17, 2023, from https://www.nytimes.com/2008/10/05/books/chapters/chap-nothing-to-be-frightened-of.html

Barrense-Dias, Y., Chok, L., & Surís, J.-C. (2021). A picture of the mental health of adolescents in Switzerland and Liechtenstein. In *Unisanté - Centre universitaire de médecine générale et santé publiqu* (pp. 1–71). https://doi.org/10.16908/issn.1660-7104/323

Blickhan, D. (2017). *Anleitung zum Glücklichsein: Studie über die Wirkung von Ausbildungen in Angewandter Positiver Psychologie auf Wohlbefinden und Flourishing der Teilnehmenden*. Freie Universität.

Boltanski, L., & Chiapello, È. (2003). *Der neue geist des kapitalismus*. BoD–Books on Demand.

Boyle, P. A., Barnes, L. L., Buchman, A. S., & Bennett, D. A. (2009). Purpose in life is associated with mortality among community-dwelling older persons. *Psychosomatic medicine*, *71*(5), 574–579. https://doi.org/10.1097/PSY.0b013e3181a5a7c0

Bröckling, U. (2013). *Das unternehmerische Selbst: Soziologie einer Subjektivierungsform*. Suhrkamp Verlag.

Bundesamt für Statistik. (2018). Gesundheit der Studierenden an den Schweizer Hochschulen. Retrieved May 17, 2023, from https://dam-api.bfs.admin.ch/hub/api/dam/assets/6526111/master

Butler, J., & Kern, M. L. (2016). The PERMA-Profiler: A brief multidimensional measure of flourishing. *International Journal of Wellbeing*, *6*(3), 1–48. https://doi.org/10.5502/ijw.v6i3.526

Campbell, J. (1949). *The Hero with a Thousand Faces*. New World Library.

Campbell, J. (1953). *Der Heros in tausend Gestalten*. Suhrkamp Verlag.

Campbell, J. (1972). *Myths to live by*. Viking Press.

Campbell, J. (1988). *The power of myth*. Anchor Books.

Campbell, J. (2004). *Pathways to bliss: Mythology and personal transformation (Bd. 16*. New World Library.

Csikszentmihalhi, M. (2020). *Finding flow: The psychology of engagement with everyday life*. Hachette Book Group.

Das optimierte Selbst. (2014). *NIM - Nürnberg Institut für Marktentscheidungen*. Retrieved March 24, 2020, from https://www.gfk-verein.org/compact/fokusthemen/das-optimierte-selbst%EF%83%B1

Diekmann, A. (2007). *Empirische Sozialforschung. Grundlagen, Methoden* (Anwendungen (18. Aufl). Rowohlt Taschenbuch.

Diener, E., & Seligman, M. E. (2002). Very happy people. *Psychological science*, *13*(1), 81–84. https://doi.org/10.1111/1467-9280.00415

Diener, E., & Seligman, M. E. (2004). Beyond money: Toward an economy of well-being. *Psychological science in the public interest*, *5*(1), 1–31. https://doi.org/10.1111/j.0963-7214.2004.00501001.x

Donaldson, J. F., Graham, S. W., Martindill, W., & Bradley, S. (2000). Adult undergraduate students: How do they define their experiences and their success? *The Journal of Continuing Higher Education*, *48*(2), 2–11. https://doi.org/10.1080/07377366.2000.10400400

Duden. (2023a). Held. Retrieved May 17, 2023, from https://www.duden.de/rechtschreibung/Held_Held_Recke

Duden. (2023b). Populismus. Retrieved May 17, 2023, from https://www.duden.de/rechtschreibung/Populismus

Duden. (2023c). Selbstoptimierung. Retrieved May 17, 2023, from https://www.duden.de/rechtschreibung/Selbstoptimierung

Duttweiler, S. (2013). *Beratung und Therapie in Zeiten des Neoliberalismus–oder: Vom Zwang zur Freiheit, sich selbst zu optimieren*. Psychotherapie & Sozialwissenschaft.

Duttweiler, S. (2016). Alltägliche (Selbst) Optimierung in neoliberalen Gesellschaften. *Aus Politik und Zeitgeschichte*, *66*(37–38), 27–32.

Ehrenberg, A. (2015). *Das erschöpfte Selbst: Depression und Gesellschaft in der Gegenwart*. Campus Verlag.

Fenner, D. (2020). Selbstoptimierung. bpb Bundeszentrale für politische Bildung. Retrieved May 17, 2023, from https://www.bpb.de/themen/umwelt/bioethik/311818/selbstoptimierung/

Flick, U., & Kardorff, E. v. (2006). *Qualitative Forschung* (I. Steinke, Ed.). Aufl.). Rowohlt Taschenbuch Verlag.

Foucault, M. (1977). Der Wille zum Wissen, Sexualität und Wahrheit.

Foucault, M. (1988). *Technologies of the self: A seminar with Michel Foucault* (L. Martin, H. Gutman, & P. Hutton, Eds.). University of Massachusetts Press.

Foucault, M. (2005). Die Ethik der Sorge um sich als Praxis der Freiheit. In D. Defert & F. Ewald (Eds.), *Michel Foucault. Schriften in vier Bänden. Dits et Ecrits. Band I* (pp. 875–902). Suhrkamp.

Foucault, M. (2015). Geschichte der Geouvernementalität II. In *Die Geburt der Biopolitik* (4. Aufl). Suhrkamp Verlag.

Franzen, A. (2019). Antwortskalen in standardisierten Befragungen.

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American psychologist*, *56*(3), 218–226. https://doi.org/10.1037/0003-066X.56.3.218

Gamm, G. (2013). Das Selbst und sein Optimum. Selbstverbesserung als das letzte Anliegen der modernen Kultur. In R. Mayer, C. Thompson, & M. Wimmer (Eds.), *Inszenierung und Optimierung des Selbs* (pp. 31–53). Springer.

Girkinger, M. (2012). *Einmal Glück und Erfolg, bitte!: Über das Glück und seine Vermarktung in der Persönlichkeitsbildung*. Eine Untersuchung zur Kultur der Selbstoptimierung. Tectum Wissenschaftsverlag.

Girkinger, M. (2019). Warum wollen wir uns selbstoptimieren? Retrieved May 17, 2023, from https://jungk-bibliothek.org/wp-

Haidt, J. (2006). *The happiness hypothesis: Finding modern truth in ancient wisdom*. Basic books.

Hainey, M. (2018). Beneath the Surface of Bruce Springsteen. *Esquire*. Retrieved May 17, 2023, from https://www.esquire.com/entertainment/a25133821/bruce-springsteen-interview-netflix-broadway-2018/

Hannich, F., & Berger, V. (2011). Schweizer Wohlstand zwischen Vergangenheit und Zukunft. https://doi.org/10.21256/zhaw-968

Herold, T. (2020). Die Heldenreise—Das Abenteuer der Selbstentdeckung. Retrieved May 17, 2023, from https://thomasherold.com/heldenreise-bestimmung/

Heßdörfer, F. (2018). Die Leere der Optimierung–Versuch über eine Ideologie ohne Idee. *Politische Psychologie*, *1*, 110–125.

Hirschman, A. O. (1977). *The passions and the interests*. Princeton University Press.

Hondo, D. (2013). Charles Taylor, exclusive humanism, and the Dharma. No Zen in the West. Retrieved May 17, 2023, from https://nozeninthewest.wordpress.com/2013/06/16/charles-taylor-exclusive-humanism-and-the-dharma/

Horkheimer, M., & Adorno, T. W. (1987). Dialektik der Aufklärung. In G. Noerr (Ed.), *Max Horkheimer: Gesammelte Schriften (B* (pp. 453–457, Vol. 5). Fischer.

Immer, N. (2017). PERMA-Profiler (Institut für Positive Psychologie und Mental Coaching. Retrieved May 17, 2023, from https://www.ippm.at/wp-content/uploads/2019/11/PERMA_Profiler_Druckversion.pdf

Jakubasch, L. (2017). Aktueller Stand der Lebenszufriedenheit von Studierenden: Eine empirische Untersuchung am Hochschulstandort Mittweida [Hochschule Mittweida.].

Jochum, G. (2010). *Arbeit als Grundlage menschlicher Existenz: Zur historischen Entwicklung des Verständnisses von Arbeit* (F. Böhle, G. Voss, G. Wachtler, & H. Arbeitssoziologie, Eds.). VS Verlag für Sozialwissenschaften.

Jungk, R. (1993). *Trotzdem: Mein Leben für die Zukunft*. Carl Hanser Verlag.

Jurt, J. (2012). Bourdieus Kapital-Theorie. In M. Bergman (Ed.), *Bildung, Arbeit, Erwachsenwerden: Ein interdisziplinärer Blick auf die Transition im Jugend- und jungen Erwachsenenalte* (pp. 21–41). Springer.

Kinsella, E. L. (2013). A psychological perspective on the features and functions of heroes. *University of Limerick (Ireland)*.

Kolmar, M. (2020). Narrative Lost Theorie [Vorlesungsfolien].

Kolmar, M. (2021). *Grenzbeschreitungen: Vom Sinn, dem gelingenden Leben und unserem Umgang mit Natur*. Vandenhoeck & Ruprecht.

Kvale, S. (1994). Ten standard objections to qualitative research interviews. *Journal of phenomenological psychology*, *25*(2), 147–173. https://doi.org/10.1163/156916294X00016

Lacan, J. (1977). *Écrits: A Selection.(trans. Alan Sheridan)*. Tavistock/Routledge.

Lacan, J. (1992). *The ethics of psychoanalysis: The seminar of Jacques Lacan*. Tavistock/Routledge.

Lengfeld, H. (2019). Abstiegsangst in Deutschland auf historischem Tiefstand: Ergebnisse der Auswertung des Sozio-oekonomischen Panels 1991–2016. In J. Delhey (Ed.), *Diagnose Angstgesellschaft?: Was wir wirklich über die Gefühlslage der Menschen wisse* (pp. 59–75). transcript Verlag.

Maio, G. (2015). „Wir leben in einer Gesellschaft, die nur Sieger sehen will". *NZZ. [Interview]*. Retrieved May 17, 2023, from https://www.nzz.ch/wissenschaft/bildung/wir-leben-in-einer-gesellschaft-die-nur-sieger-sehen-will-ld.732406

Maslow, A. H. (2013). *Toward a psychology of being*. Simon and Schuster.

Mayring, P. (1991). Qualitative Inhaltsanalyse. In U. Flick, E. Kardoff, H. Keupp, L. Rosenstiel, & W. S (Eds.), *Handbuch qualitative Forschung: Grundlagen, Konzepte, Methoden und Anwendunge* (pp. 209–213). Beltz - Psychologie Verlags Union.

Mecke, I. (2001). Das narrative Interview. Retrieved May 17, 2023, from http://webdoc.sub.gwdg.de/ebook/diss/2003/fu-berlin/2001/224/Kap.6_Teil.I.VI.pdf

Medick, V. (2015). Islamischer Staat: „Der IS ist für viele schlicht ein Abenteuer". *Der Spiegel*. Retrieved May 17, 2023, from https://www.spiegel.de/politik/ausland/islamischer-staat-der-is-ist-fuer-viele-schlicht-ein-abenteuer-a-1065754.html

Meissner, S. (2016). Der vermessene Schlaf. In S. Duttweiler, R. Gugutzer, J.-H. Passoth, & J. Strübing (Eds.), *Leben nach Zahlen: Self-Tracking als Optimierungsprojekt* (p. 10). transcript Verlag.

Moors, G. (2008). Exploring the effect of a middle response category on response style in attitude measurement. *Quality & quantity*, *42*, 779–794. https://doi.org/10.1007/s11135-006-9067-x

Mühlhausen, C. (2016a). Healthstyle III: Ganzheitlich gesund! Aus Selbstoptimierung wird Self Coaching. Retrieved May 17, 2023, from https://docplayer.org/54709909-Healthstyle-iii-ganzheitlich-gesund-aus-selbstoptimierung-wird-self-coaching.html

Mühlhausen, C. (2016b). Selbstoptimierung 2.0—Entspannt schneller, höher und weiter kommen. na-news aktuell. Retrieved May 17, 2023, from https://docplayer.org/35945132-Whitepaper-selbstoptimierung-selbstoptimierung-wie-menschen-und-unternehmen-vom-megatrend-profitieren.html

Mühlhausen, C., & Wippermann, P. (2013). *Das Zeitalter der Selbstoptimierer*. New Business Verlag.

Münkler, H. (2011). *Die neuen Kriege*. Rowohlt Verlag GmbH.

Münkler, H. (2015). *Kriegssplitter: Die Evolution der Gewalt im 20. Und 21*. Jahrhundert. Rowohlt Verlag GmbH.

Nachtwey, O. (2016). *Die Abstiegsgesellschaft: Über das Aufbegehren in der regressiven Moderne*. Suhrkamp Verlag.

Neiman, S. (2010). Victims and Heroes: The Tanner Lecture on Human Values. Retrieved May 17, 2023, from https://deepblue.lib.umich.edu/bitstream/handle/2027.42/66469/neiman-tanner_lecture.mp4?sequence=1

Neumüller, D. (2013). *Reaktionen auf soziale Ausgrenzung in kollektivistischen und individualistischen Kulturen*. Universität Graz.

Perriard, J., & Ramaciotti, D. (2003). Die Kosten des Stresses in der Schweiz. *Staatssekretariat für Wirtschaft*. Retrieved May 17, 2023, from https://www.seco.admin.ch/seco/de/home/Publikationen_Dienstleistungen/Publikationen_und_Formulare/Arbeit/Arbeitsbedingungen/Studien_und_Berichte/die-kosten-des-stresses-in-der-schweiz.html

Peterson, C., & Seligman, M. E. (2012). Character Strengths and Virtues: A Handbook and Classification.

Picton, C., Kahu, E. R., & Nelson, K. (2018). Hardworking, determined and happy': First-year students' understanding and experience of success. *Higher education research & development*, *37*(6), 1260–1273. https://doi.org/10.1080/07294360.2018.1478803

Questionnaires | Moral Foundations Theory. (n.d.). Retrieved May 17, 2023, from https://moralfoundations.org/questionnaires/

Rebentisch, J. (2012). Hegels Missverständnis der ästhetischen Freiheit. In J. Rebentisch & C. Menke (Eds.), *Kreation und Depression: Freiheit im gegenwärtigen Kapitalismu* (2. Aufl., Bd, pp. 172–190, Vol. 67). Kulturverlag Kadmos.

Reckwitz, A. (2017). *Die Gesellschaft der Singularitäten (1*. Aufl.). Suhrkamp Verlag.

Response-Bias: Verzerrung von Ergebnissen durch Teilnehmer - Wirtschaftspsychologische Gesellschaft. (n.d.). Retrieved May 18, 2023, from https://wpgs.de/fachtexte/ergebnisinterpretation/response-bias-verzerrung-ergebnisse-durch-teilnehmer/

Röcke, A. (2021). *Soziologie der Selbstoptimierung*. Suhrkamp Verlag.

Rosa, H. (2005). *Beschleunigung: Die Veränderung der Zeitstrukturen in der Moderne*. Suhrkamp Verlag.

Rosa, H. (2019). Ich will den Modus unseres In-der-Welt-Seins ändern. *Philosophie Magazin*. Retrieved May 17, 2023, from https://www.philomag.de/artikel/hartmut-rosa-ich-will-den-modus-unseres-der-welt-seins-aendern

Rose, N. (1999). *Governing the soul: The shaping of the private self* (2. Aufl). Free Association Books.

Ryff, C. D., Singer, B. H., & Dienberg Love, G. (2004). Positive health: Connecting well–being with biology. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *359*(1449), 1383–1394. https://doi.org/10.1098/rstb.2004.1521

Salazar, M. K. (1990). Interviewer Bias: How it Affects Survey Research. *AAOHN Journal*, *38*(12), 567–572. https://doi.org/10.1177/216507999003801203

Schmelzer, T. (2015). Sei authentisch! Von wegen. *Der Standard*. Retrieved May 17, 2023, from http://www.pressreader.com/austria/der-standard/20190525/282647508990439

Schramek, G. G. R. (2012). Angst, Depressivität und Lebenszufriedenheit von Studentinnen und Studenten der Universität Leipzig.

Seligman, M. E. (2002). *Authentic happiness: Using the new positive psychology to realize your potential for lasting fulfillment*. Simon; Schuster.

Seligman, M. E. (2018). PERMA and the building blocks of well-being. *The Journal of Positive Psychology*, *13*(4), 333–335. https://doi.org/10.1080/17439760.2018.1437466

Smith, B. W., Ford, C. G., Erickson, K., & Guzman, A. (2021). The Effects of a Character Strength Focused Positive Psychology Course on Undergraduate Happiness and Well-Being. *Journal of Happiness Studies*, *22*(1), 343–362. https://doi.org/10.1007/s10902-020-00233-9

Smith, J. (2019). Charles Taylor and Our Secular Age. In *Southeastern Baptist Theological Seminary*. Retrieved May 17, 2023, from https://www.youtube.com/watch?v=H9Ka32RdcTM

Sprenger, M. (2017). An sich arbeiten – Zum Zwang zur Selbstinvestition. *Universität St. Gallen*.

Steger, M. F. (2013). Experiencing meaning in life: Optimal functioning at the nexus of well-being, psychopathology, and spirituality. In *The human quest for meanin* (pp. 211–230). Routledge.

Taylor, C. (2007). *A secular age*. Harvard University Press.

Taylor, S. E. (2011). Social support: A review. In H. Friedman (Ed.), *The Oxford handbook of health psycholog* (pp. 189–214). Oxford University Press.

UCLA CTSI Community Engagement & Research Program. (n.d.). How to: Pretest a survey questionnaire. Retrieved May 17, 2023, from https://ctsi.ucla.edu/patients-community/files/view/docs/Howto_pretest_survey_questionnaire.pdf

Ullrich, W. (2016). *Habenwollen: Wie funktioniert die Konsumkultur? S*. Fischer Verlag.

Venter, H. J. (2016). Self-transcendence: Maslow's answer to cultural closeness. *Journal of Innovation Management*, *4*(4), 3–7. https://doi.org/10.24840/2183-0606_004.004_0002

Vormbusch, U. (2016). Taxonomien des Selbst. In S. Duttweiler (Ed.), *Leben nach Zahlen. Self-Tracking als Optimierungsprojek* (pp. 45–59). transcript Verlag.

Weber, M. (2002). *Wirtschaft und Gesellschaft: Grundriss der Verstehenden Soziologie (5*. Aufl.). JCB Mohr.

Weber, M. (2016). Die protestantische Ethik und der „Geist" des Kapitalismus. In K. Lichtblau & J. Weiss (Eds.), *Die protestantische Ethik und der „Geist" des Kapitalismus (B* (pp. 29–176, Vol. 53). Springer Fachmedien.

Williams, C. L., & Heikes, E. J. (1993). The importance of researcher's gender in the in-depth interview: Evidence from two case studies of male nurses. *Gender & society*, *7*(2), 280–291. https://doi.org/10.1177/089124393007002008

Zeidner, M. (2021). „Don't worry—Be happy": The sad state of happiness research in gifted students. *High Ability Studies*, *32*(2), 125–142. https://doi.org/10.1080/13598139.2020.1733392